Single-Look Multi-Master SAR Tomography: An Introduction

Nan Ge, Richard Bamler, *Fellow, IEEE*, Danfeng Hong, *Member, IEEE*, and Xiao Xiang Zhu, *Senior Member, IEEE*

***Abstract*— This article addresses the general problem of single-look multi-master SAR tomography. For this purpose, we establish the single-look multi-master data model, analyze its implications for the single and double scatterers, and propose a generic inversion framework. The core of this framework is the nonconvex sparse recovery, for which we develop two algorithms: one extends the conventional nonlinear least squares (NLS) to the single-look multi-master data model and the other is based on bi-convex relaxation and alternating minimization (BiCRAM). We provide two theorems for the objective function of the NLS subproblem, which lead to its analytic solution up to a constant phase angle in the 1-D case. We also report our findings from the experiments on different acceleration techniques for BiCRAM. The proposed algorithms are applied to a real TerraSAR-X data set and validated with the height ground truth made available by an SAR imaging geodesy and simulation framework. This shows empirically that the *single-master* approach, if applied to a single-look *multi-master* stack, can be insufficient for layover separation, and the *multi-master* approach can indeed perform slightly better (despite being computationally more expensive) even in the case of single scatterers. In addition, this article also sheds light on the special case of single-look bistatic SAR tomography, which is relevant for the current and future SAR missions such as TanDEM-X and Tandem-L.**

***Index Terms*— Bistatic SAR, nonconvex optimization, SAR tomography, sparse recovery, synthetic aperture radar (SAR), Tandem-L, TanDEM-X.**

I. Introduction

SYNTHETIC APERTURE RADAR (SAR) tomography is an interferometric SAR (InSAR) technique that reconstructs a 3-D far field from the 2-D azimuth-range measurements of the radar echoes [1]–[3]. In the common case of spaceborne repeat-pass acquisitions, the scatterers' motion can also be modeled and estimated [4]–[6]. SAR tomography is sometimes considered as an extension of persistent scatterer interferometry (PSI) [7]–[9] to the multi-scatterer case, although the inversion of the latter is performed on the double-difference phase observations of the persistent scatterers (PS) [10]. Extensive efforts were devoted to improving the super-resolution power, robustness, and computational efficiency of the tomographic inversion in the urban scenarios [11]–[20].

The publications on SAR tomography can be roughly classified into the following four categories (see also Table I). Note that those listed below were only handpicked, and we have no intention to provide a complete list.

1) *Single-Look Single-Master:* Reigber and Moreira [1] did the pioneering work on airborne SAR tomography by densifying sampling through the integer interferogram-combination technique and subsequently employing discrete Fourier transform on an interpolated linear array of baselines. Fornaro *et al.* [3], [5], [21] paved the way for spaceborne SAR tomography with long-term repeat-pass acquisitions and proposed to use more advanced inversion techniques such as truncated singular value decomposition. Zhu and Bamler [22] provided the first demonstration of SAR tomography with very high-resolution spaceborne SAR data by using Tikhonov regularization and nonlinear least squares (NLS). Zhu and Bamler [6] and Budillon *et al.* [11] introduced compressive-sensing techniques to tomographic inversion under the assumption of a compressible far-field profile. Zhu and Bamler [23] proposed a generic algorithm (named SL1MMER) that is composed of spectral estimation, model-order selection, and debiasing.
2) *Single-Look Multi-Master*[1]*:* To the best of our knowledge, the publications in this category are rather scarce. Zhu and Bamler [24] extended the Tikhonov regularization, NLS, and compressive-sensing approaches to a mixed TerraSAR-X and TanDEM-X stack by using the preestimated covariance matrix. Ge and Zhu [25] proposed a framework for SAR tomography using only the bistatic or pursuit monostatic acquisitions: nondifferential SAR tomography for height estimation by using

Manuscript received November 1, 2019; revised March 9, 2020 and May 2, 2020; accepted June 6, 2020. This work was supported in part by the European Research Council (ERC) through the European Union's Horizon 2020 Research and Innovation Programme (*So2Sat*) under Grant ERC-2016-StG-714087 and in part by the Helmholtz Association through the Framework of Helmholtz Artificial Intelligence Cooperation Unit (HAICU)—Local Unit "Munich Unit @Aeronautics, Space and Transport (MASTr)" and Helmholtz Excellent Professorship "Data Science in Earth Observation—Big Data Fusion for Urban Research." *(Corresponding author: Xiao Xiang Zhu.)*

Nan Ge and Danfeng Hong are with Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Wessling, Germany (e-mail: nan.ge@dlr.de; danfeng.hong@dlr.de).

Richard Bamler is with Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Wessling, Germany, and also with the Chair of Remote Sensing Technology, Technical University of Munich (TUM), 80333 Munich, Germany (e-mail: richard.bamler@dlr.de).

Xiao Xiang Zhu is with Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Wessling, Germany, and also with the Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), 80333 Munich, Germany (e-mail: xiaoxiang.zhu@dlr.de).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2020.3002945

[1]In this context, "multi-master" can be interpreted as "not single-master" (see also our definition in Section II-A).

the bistatic or pursuit monostatic interferograms and differential SAR tomography for deformation estimation by using conventional repeat-pass interferograms and the previous height estimates as the deterministic prior [25]. However, the single-look single-master data model still underlies the algorithms in both publications.

3) *Multi-Look Single-Master:* Aguilera *et al.* [13] exploited the common sparsity pattern among multiple polarimetric channels by distributed compressive sensing. Schmitt and Stilla [14] also employed distributed compressive sensing to reconstruct jointly an adaptively chosen neighborhood. Liang *et al.* [26] proposed an algorithm for 2-D range-elevation focusing on azimuth lines by compressive sensing. Shi *et al.* [20] performed nonlocal InSAR filtering before tomographic reconstruction.
4) *Multi-Look Multi-Master:* In general, any algorithm estimating the autocorrelation matrix belongs to this category. Note that this is closely related to the modern adaptive multi-looking techniques that also exploit all possible interferometric combinations [27]–[31]. Gini *et al.* [2] investigated the performance of different spectral estimators including Capon, multiple signal classification (MUSIC), and the multi-look relaxation (M-RELAX) algorithm. Lombardini [4] extended SAR tomography to the differential case by formulating it as a multi-dimensional spectral estimation problem and tackled it with a higher-order Capon. Duque *et al.* [32], [33] were the first to investigate bistatic SAR tomography by using ground-based receivers and spectral estimators such as Capon and MUSIC. Duque *et al.* [34] demonstrated the feasibility of SAR tomography using a single pass of alternating bistatic acquisitions, in which the eigendecomposed empirical covariance matrix was exploited for the hypothesis test on the number of scatterers [34]. Fornaro *et al.* [15] proposed an algorithm (named CAESAR) employing principal component analysis of the eigendecomposed empirical covariance matrix in an adaptively chosen neighborhood.

This list has a clear focus on the urban scenarios. Needless to say, SAR tomography in the forested scenarios involving random volume scattering in the canopy and the double-bounce scattering between the ground and the trunk [35]–[42] also falls in the *multi-look multi-master* category.

Let us follow the common conventions and denote the azimuth, range, and elevation axes as x, r, and s, respectively, where s is perpendicular to the xr plane. For the sake of argument, suppose for any sample at the x and r positions, the N single-look complex (SLC) SAR measurements are noiseless. After deramping, each phase-calibrated SLC measurement can be modeled as the Fourier transform Γ of the elevation-dependent far-field reflectivity function $\gamma : \mathbb{R} \to \mathbb{C}$ at the corresponding wavenumber k [21]

$$g_n = \Gamma(k_n) \coloneqq \int \gamma(s) \exp(-jk_n s) \mathrm{d}s, \quad n = 1, \ldots, N \quad (1)$$

where $k_n \coloneqq -4\pi b_n/(\lambda r_0)$ is the nth wavenumber determined by the sensor position b_n along an axis $b \parallel s$ with respect to an arbitrary reference, the radar wavelength λ, and the slant-range distance r_0 with respect to a ground reference point. Here, we consider the nondifferential case. An extension to the differential case, in which the scatterers' motion is modeled as the linear combination of the basis functions, is straightforward.

In the *single-look single-master* case, one SLC (say the ith, $i \in [N]$), typically near the center of the joint orbital and temporal distribution, is selected as the unique master for generating interferograms, i.e., $g_n\overline{g_i}/|g_i|$, $\forall n \in [N] \setminus \{i\}$. This process that can also be interpreted as a phase-calibration step converts k_n into the wavenumber baseline $\Delta k_n \coloneqq k_n - k_i$, $\forall n \in [N]$. As a result, the zero position of the wavenumber baseline is fixed, i.e., $\Delta k_i = 0$. The rationale behind this is, for example, to facilitate 2-D phase unwrapping for atmospheric-phase-screen (APS) compensation by smoothing out the interferometric phase in xr.

Likewise, the data model of random volume scattering is straightforward in the *multi-look multi-master* setting. Suppose that $\gamma(s)$ is a white random signal. For any master and slave sampled at k and $k+\Delta k$, respectively, the Van Cittert–Zernike theorem implies that the expectation (due to multi-looking) of the interferogram, being the autocorrelation function $R_{\Gamma\Gamma}$ of Γ, is the Fourier transform of the elevation-dependent backscatter coefficient function $\sigma_0 : \mathbb{R} \to \mathbb{R}$ at Δk

$$\mathrm{E}\left[\Gamma(k+\Delta k)\overline{\Gamma(k)}\right] = R_{\Gamma\Gamma}(\Delta k) = \int \sigma_0(s) \exp(-j\Delta k s) \mathrm{d}s \quad (2)$$

where the property of $\gamma(s)$ being white, i.e., $\mathrm{E}[\gamma(s)\overline{\gamma(s')}] = \sigma_0(s)\delta(s-s')$, is used. This leads to an inverse problem similar to the one in the *single-look single-master* case.

This article, on the other hand, addresses the general problem of SAR tomography using a *single-look multi-master* stack. Such a stack arises when, for example: 1) a stack of bistatic interferograms is used in order to diminish the APS, to minimize the temporal decorrelation of the non-PSs, and to eliminate the motion-induced phase of the single scatterers [33], [43], [44] and 2) the repeat-pass interferograms of small (temporal) baselines are employed so as to limit the corresponding decorrelation effects of the non-PSs [45]–[47]. While both previously mentioned categories have been intensively studied, it is not the case for *single-look multi-master* SAR tomography. To the best of our knowledge, all the existing work to date toward single-look *multi-master* SAR tomography is still incorrectly based on the single-look *single-master* data model [24], [25]. As will be demonstrated later with a real SAR data set, this approach can be insufficient for layover separation, even if the elevation distance between two scatterers is significantly larger than the Rayleigh resolution. This motivates us to fill the gap in the literature by revisiting the single-look data model in a multi-master multi-scatterer configuration and by developing efficient methods for tomographic reconstruction. Naturally, this study is also inspired by the prospective SAR missions such as Tandem-L that will deliver high-resolution wide-swath bistatic acquisitions in the L-band as the operational products [48].

Our main contributions can be summarized as follows.

TABLE I
CLASSIFICATION OF TOMOGRAPHIC SAR ALGORITHMS

	Single-Master	Multi-Master
Single-Look	Reigber & Moreira (2000) Fornaro et al. (2003, 2005, 2008) Budillon et al. (2010) Zhu & Bamler (2010a, 2010b, 2011) Etc.	Zhu & Bamler (2012)† Ge & Zhu (2019)†
Multi-Look	Aguilera et al. (2012) Schmitt & Stilla (2012) Liang et al. (2018) Shi et al. (2019) Etc.	Gini et al. (2002) Lombardini (2005) Duque et al. (2009, 2010, 2014) Fornaro et al. (2014) Etc.

† Incorrectly uses the single-look single-master data model

1) We establish the data model of single-look multi-master SAR tomography by means of which both sparse recovery and model-order selection can be formulated as the nonconvex minimization problems.
2) We develop two algorithms for solving the aforementioned nonconvex sparse recovery problem.
   a) *NLS:* We provide two theorems regarding the critical points of its subproblem's objective function that also underlies model-order selection.
   b) *Bi-Convex Relaxation and Alternating Minimization (BiCRAM):* We propose to sample its solution path for the purpose of automatic regularization parameter tuning, and we show empirically that a simple diagonal preconditioning can effectively improve convergence.
3) We propose to correct the quantization errors by using (nonconvex) nonlinear optimization.
4) We validate the tomographic height estimates with the ground truth generated by SAR simulations and geodetic corrections.

The rest of this article is organized as follows. Section II introduces the data model and inversion framework for the single-look multi-master SAR tomography. In Sections III and IV, two algorithms for solving the nonconvex sparse recovery problem within the aforementioned framework, namely, NLS and BiCRAM, are elucidated and analyzed, respectively. Section V reports an experiment with the TerraSAR-X data including a validation of the tomographic height estimates. This article is concluded by Section VI.

II. SINGLE-LOOK MULTI-MASTER SAR TOMOGRAPHY

In this section, we establish the data model for single-look multi-master SAR tomography, analyze its implications for two specific cases, and sketch out a generic inversion framework for it.

We start with the mathematical notations that are used throughout this article.

Notation: We denote scalars as lower- or uppercase letters (e.g., m, N, and λ), vectors as bold lowercase letters (e.g., $\mathbf{g}$ and $\boldsymbol{\gamma}$), matrices, sets, and ordered pairs as bold uppercase letters (e.g., $\mathbf{R}$ and $\boldsymbol{\Omega}$), and the number fields as blackboard bold uppercase letters (e.g., $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$) with the following conventions:

1) g_n denotes the nth entry of $\mathbf{g}$.
2) $\mathbf{a}^m$ and $\mathbf{a}_n$ denote the mth row and nth column of $\mathbf{A}$, respectively.
3) $\mathrm{Diag}(\mathbf{a})$ denotes a square diagonal matrix whose entries on the main diagonal are equal to $\mathbf{a}$ and $\mathrm{Diag}(\mathbf{A})$ denotes a vector whose entries are equal to those on the main diagonal of $\mathbf{A}$.
4) $\mathrm{Supp}(\mathbf{x})$ denotes the index set of the nonzero entries or support of $\mathbf{x}$.
5) $\overline{\mathbf{A}}$, $\mathbf{A}^T$, and $\mathbf{A}^H$ denote the (elementwise) complex conjugate, transpose, and conjugate transpose of $\mathbf{A}$, respectively.
6) $\mathbf{A}_R$ and $\Re(\mathbf{A})$ denote the real part of $\mathbf{A}$.
7) $\mathbf{A}_I$ and $\Im(\mathbf{A})$ denote the imaginary part of $\mathbf{A}$.
8) $\mathbf{A} \circ \mathbf{B}$ denotes the Hadamard product of $\mathbf{A}$ and $\mathbf{B}$.
9) $\mathbf{A} \succ \mathbf{0}$, $\mathbf{B} \prec \mathbf{0}$ means that $\mathbf{A}$ is positive-definite and $\mathbf{B}$ is negative-definite.
10) $\mathbf{A}_{\boldsymbol{\Omega}}$ denotes the matrix formed by extracting the columns of $\mathbf{A}$ indexed by $\boldsymbol{\Omega}$.
11) $\|\mathbf{A}\|_{1,2}$ denotes the $\ell_{1,2}$ norm of $\mathbf{A}$, i.e., the sum of the ℓ_2 norms of its rows.
12) $\mathbf{I}$ denotes the identity matrix.
13) $[N]$ denotes the set $\{1, \ldots, N\}$.
14) $|\boldsymbol{\Omega}|$ denotes the cardinality of the set $\boldsymbol{\Omega}$.
15) The nonnegative and positive subsets of a number field $\mathbb{F}$ are denoted as $\mathbb{F}_+$ and $\mathbb{F}_{++}$, respectively.

A. Data Model

First, we give a definition for "single-master" and "multi-master" by using the language of the basic graph theory (see [49, §1]). Let $\mathbf{G} \coloneqq (\mathbf{V}(\mathbf{G}), \mathbf{E}(\mathbf{G}))$ be an acyclic directed graph that is associated with an incidence function $\psi_{\mathbf{G}}$, where $\mathbf{V}(\mathbf{G}) \coloneqq [N]$ is a set of vertices (SLCs), $\mathbf{E}(\mathbf{G})$ is a set of edges (interferograms), and for each $e \in \mathbf{E}(\mathbf{G})$, $\exists m, n \in \mathbf{V}(\mathbf{G})$ such that $\psi_{\mathbf{G}}(e) = (m, n)$. Its adjacency matrix $\mathbf{A}(\mathbf{G}) \coloneqq (a_{m,n}) \in \{0, 1\}^{N \times N}$ is given by

$$a_{m,n} \coloneqq \begin{cases} 1 & : (m, n) \in \mathbf{E}(\mathbf{G}), \\ 0 & : (m, n) \notin \mathbf{E}(\mathbf{G}). \end{cases} \tag{3}$$

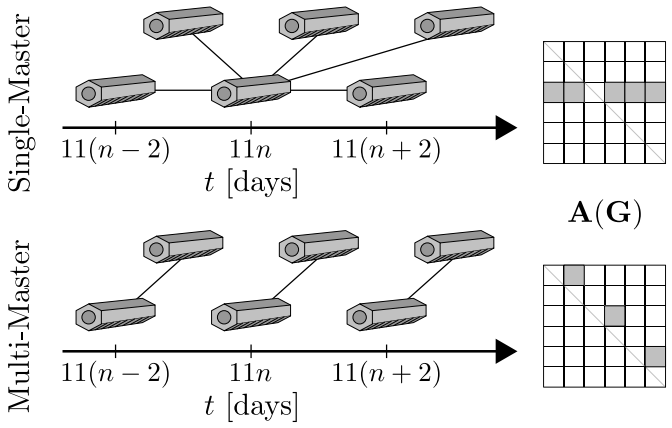


Fig. 1. Single-master versus multi-master: two exemplary configurations and the corresponding adjacency matrices $\mathbf{A}(\mathbf{G})$.

Since $\mathbf{G}$ is acyclic, $a_{n,n} = 0$, $\forall n \in \mathbf{V}(\mathbf{G})$, i.e., the diagonal of $\mathbf{A}(\mathbf{G})$ contains only zero entries. Without loss of generality, assume that every vertex is connected to at least another one.

Definition 1: The *single-master* configuration means that there exists a unique $i \in [N]$ such that $a_{i,n} = 1$, $\mathbf{a}^m = \mathbf{0}$, $\forall m, n \in [N] \setminus \{i\}$. In this case, we refer to $\{g_n\overline{g_i}/|g_i|\}$ as the *single-master* stack with a master indexed by i.

Definition 2: The *multi-master* configuration means that $\nexists i \in [N]$ such that $a_{i,n} = 1$, $\mathbf{a}^m = \mathbf{0}$, $\forall m, n \in [N] \setminus \{i\}$. In this case, we refer to $\{g_n\overline{g_m}\}$ as the *multi-master* stack.

That is, "multi-master" is equivalent to "not single-master." Two exemplary configurations are illustrated in Fig. 1.

In the *multi-master* case, an interferogram is created for each $(m, n) \in \mathbf{E}(\mathbf{G})$

$$g_n\overline{g_m} = \int\int \gamma(s)\overline{\gamma(s')}\exp\big(-j\big(k_n s - k_m s'\big)\big)\mathrm{d}s\mathrm{d}s'. \quad (4)$$

Hereafter, we focus on the case in which the far field contains only a small number of scatterers such that

$$g_n \approx \sum_l \gamma_l \exp(-jk_n s_l), \quad n = 1, \ldots, N \quad (5)$$

where $\gamma_l \in \mathbb{C}$ is the reflectivity of the lth scatterer located at the elevation position s_l. The single-look multi-master data model (4) becomes

$$g_n\overline{g_m} \approx \sum_{l,l'} \gamma_l\overline{\gamma_{l'}}\exp(-j(k_n s_l - k_m s_{l'})) \quad (6)$$

$\forall (m, n) \in \mathbf{E}(\mathbf{G})$.

In the next section, we analyze the implications of (6) for the single- and double-scatterer cases.

B. Implications

In the *single-scatterer* case, (6) becomes

$$g_n\overline{g_m} \approx |\gamma|^2\exp(-j(k_n - k_m)s) \quad (7)$$

that is, the multi-master observation is actually the Fourier transform of the reflectivity *power* at the wavenumber baseline $k_n - k_m$. As a result, the nonnegativity of $|\gamma|^2$ should be considered during inversion. Since both the real and imaginary parts of $g_n\overline{g_m}$ are parameterized by $|\gamma|^2$, that is

$$\begin{aligned}\Re(g_n\overline{g_m}) &\approx |\gamma|^2\cos((k_n - k_m)s)\\ \Im(g_n\overline{g_m}) &\approx |\gamma|^2\sin(-(k_n - k_m)s),\end{aligned} \quad (8)$$

the inversion problem can be recast as a real-valued one.

For *double scatterers*, the multi-master observation is

$$\begin{aligned}g_n\overline{g_m} \approx\ & |\gamma_1|^2\exp(-j(k_n - k_m)s_1)\\ &+ \gamma_1\overline{\gamma_2}\exp(-j(k_n s_1 - k_m s_2))\\ &+ \overline{\gamma_1}\gamma_2\exp(-j(k_n s_2 - k_m s_1))\\ &+ |\gamma_2|^2\exp(-j(k_n - k_m)s_2).\end{aligned} \quad (9)$$

In addition to the Fourier transform of the reflectivity power at $k_n - k_m$, the right-hand side of (9) contains the second and third "cross-terms," in which the reflectivity values of the two scatterers (and their frequency–time products) are coupled. This essentially rules out any linear model.

Remark: In the *multi-look* multi-master setting, the data model under random volume scattering is

$$\mathrm{E}\big[\Gamma(k_n)\overline{\Gamma(k_m)}\big] = \int \sigma_0(s)\exp(-j(k_n - k_m)s)\mathrm{d}s \quad (10)$$

as already indicated in (2), i.e., no coupling is involved.

Remark: A *multi-master* bistatic or pursuit monostatic (i.e., 10-s temporal baseline [50]) stack is in general not motion-free for the double (or multiple) scatterers.

To see this, consider, for example, the linear deformation model $d(t_n) := vt_n$, where v and t denote the linear deformation rate and the temporal baseline, respectively. Observe that

$$\begin{aligned}&g_n\overline{g_m}\\ &\approx \sum_{l,l'}\gamma_l\overline{\gamma_{l'}}\\ &\quad\cdot\exp(-j(k_n s_l - k_m s_{l'} + 4\pi d_l(t_n)/\lambda - 4\pi d_{l'}(t_m)/\lambda))\\ &= |\gamma_1|^2\exp(-j((k_n - k_m)s_1 + 4\pi v_1(t_n - t_m)/\lambda))\\ &\quad+\gamma_1\overline{\gamma_2}\exp(-j((k_n s_1 - k_m s_2) + 4\pi(v_1 t_n - v_2 t_m)/\lambda))\\ &\quad+\overline{\gamma_1}\gamma_2\exp(-j((k_n s_2 - k_m s_1) + 4\pi(v_2 t_n - v_1 t_m)/\lambda))\\ &\quad+|\gamma_2|^2\exp(-j((k_n - k_m)s_2 + 4\pi v_2(t_n - t_m)/\lambda)).\end{aligned} \quad (11)$$

In the case of $t_m = t_n$, the motion-induced phase in the cross-terms vanishes if and only if $v_1 = v_2$.

The next section introduces a generic inversion framework for the single-look multi-master SAR tomography.

C. Inversion Framework

The data model (6) already indicates a nonlinear system of equations for a single-look multi-master stack. Suppose $\mathbf{G}$ is the graph associated with this stack that contains a total of $N' := |\mathbf{E}(\mathbf{G})|$ multi-master observations and $e_1, \ldots, e_{N'}$ is an ordered sequence of all the edges in $\mathbf{E}(\mathbf{G})$. Let $\mathcal{M}, \mathcal{S} : [N'] \to [N]$ be the mappings to the master and slave image indices, respectively. For each e_n, $n \in [N']$, a multi-master observation $g_{\mathcal{S}(n)}\overline{g_{\mathcal{M}(n)}}$ is obtained. Let $\mathbf{g} \in \mathbb{C}^{N'}$ be the vector of the multi-master observations such that $g_n := g_{\mathcal{S}(n)}\overline{g_{\mathcal{M}(n)}}$, $\forall n \in [N']$. Let $s_1, \ldots, s_L$ be a discretization of the elevation axis s. The data model in the matrix notations is

$$\mathbf{g} \approx (\mathbf{R}\boldsymbol{\gamma}) \circ \big(\overline{\mathbf{S}\boldsymbol{\gamma}}\big) \quad (12)$$

where $\mathbf{R}$ and $\mathbf{S} \in \mathbb{C}^{N' \times L}$ represent the tomographic observation matrices of the slave and master images, respectively, $r_{n,l} \coloneqq \exp(-jk_{\mathcal{S}(n)}s_l)$, $s_{n,l} \coloneqq \exp(-jk_{\mathcal{M}(n)}s_l)$, $\forall n \in [N']$, $l \in [L]$, and $\boldsymbol{\gamma} \in \mathbb{C}^L$ is the unknown reflectivity vector such that γ_l is associated with the scatterer (if any) at the elevation position s_l.

In light of (12), we propose the following framework for tomographic inversion.

1) Nonconvex Sparse Recovery: We consider the problem

$$\begin{aligned} \hat{\boldsymbol{\gamma}} \coloneqq \quad & \arg\min_{\boldsymbol{\gamma}} \ \frac{1}{2}\|(\mathbf{R}\boldsymbol{\gamma}) \circ \left(\overline{\mathbf{S}\boldsymbol{\gamma}}\right) - \mathbf{g}\|_2^2 \\ & \text{s.t.} \ \ |\operatorname{Supp}(\boldsymbol{\gamma})| \leq K \end{aligned} \tag{13}$$

where $K \in \mathbb{Z}_{++}$. The objective function measures the model goodness of fit and the constraint enforces $\boldsymbol{\gamma}$ to be sparse, as is implicitly assumed in (6). If $\sum_{l=0}^{K} \binom{L}{l}$ is small, (13) can be solved heuristically by using the algorithms that will be developed in Section III. Section IV is dedicated to another algorithm that solves a similar problem based on bi-convex relaxation.

2) Model-Order Selection: This procedure removes the outliers and, therefore, reduces the false-positive rate. By using, for example, the Bayesian information criterion, model-order selection can be formulated as the following constrained minimization problem:

$$\begin{aligned} \hat{\boldsymbol{\Omega}} \coloneqq \quad & \arg\min_{\boldsymbol{\Omega}(,\boldsymbol{\delta})} \ 2\ln\left(\|(\mathbf{R}_{\boldsymbol{\Omega}}\boldsymbol{\delta}_{\boldsymbol{\Omega}}) \circ \left(\overline{\mathbf{S}_{\boldsymbol{\Omega}}\boldsymbol{\delta}_{\boldsymbol{\Omega}}}\right) - \mathbf{g}\|_2^2/N'\right) \\ & \qquad + (5|\boldsymbol{\Omega}| + 1)\ln\left(N'\right)/N' \\ & \text{s.t.} \ \ \operatorname{Supp}(\boldsymbol{\delta}) = \boldsymbol{\Omega} \subset \operatorname{Supp}(\hat{\boldsymbol{\gamma}}) \end{aligned} \tag{14}$$

where $\boldsymbol{\delta} \in \mathbb{C}^L$ is an auxiliary variable and $\boldsymbol{\Omega}$ is its support. Since $|\operatorname{Supp}(\hat{\boldsymbol{\gamma}})|$ is typically small, (14) can be tackled by solving a sequence of subset NLS problems in the form of

$$\min_{\boldsymbol{\epsilon}} \frac{1}{2}\|(\mathbf{R}_{\boldsymbol{\Omega}}\boldsymbol{\epsilon}) \circ \left(\overline{\mathbf{S}_{\boldsymbol{\Omega}}\boldsymbol{\epsilon}}\right) - \mathbf{g}\|_2^2 \tag{15}$$

for which two solvers will be introduced in Section III.

3) Off-Grid Correction: The off-grid or quantization problem arises when the scatterers are not located on the (regular) grid of the discrete elevation positions $s_1, \ldots, s_L$. Ge *et al.* [17] proposed to oversample $\hat{\boldsymbol{\gamma}}$ in the vicinity of the selected scatterers in order to circumvent this problem. Here, we propose a more elegant approach that is based on nonlinear optimization.

Denote $\hat{K} \coloneqq |\hat{\boldsymbol{\Omega}}|$ as the number of scatterers after model-order selection. Let ${\gamma_l}^R$ and ${\gamma_l}^I$ be the real and imaginary parts of the complex-valued reflectivity γ_l of the lth scatterer that is located at s_l, respectively, i.e., $\gamma_l = {\gamma_l}^R + j{\gamma_l}^I$, $\forall l \in [\hat{K}]$. On the basis of the single-look multi-master data model (6), we seek a solution of the following minimization problem:

$$\begin{aligned} \min_{{\gamma_l}^R, {\gamma_l}^I, s_l} \sum_n \Bigg| \ g_n - \sum_{l,l'} \left({\gamma_l}^R + j{\gamma_l}^I\right)\left({\gamma_{l'}}^R - j{\gamma_{l'}}^I\right) \\ \cdot \exp\left(-j\left(k_{\mathcal{S}(n)}s_l - k_{\mathcal{M}(n)}s_{l'}\right)\right)\Bigg|^2. \end{aligned} \tag{16}$$

Note that the objective function is differentiable with respect to ${\gamma_l}^R$, ${\gamma_l}^I$, and s_l, $\forall l \in [\hat{K}]$. Needless to say, the on-grid estimates from (14) are used as the initial solution. We will revisit this problem in Section III-A.

Thus far, the inversion framework has been established. In the next two sections, we will deal with the optimization problems (13)–(16) from the algorithmic point of view.

III. Nonlinear Least Squares

NLS is a parametric method that breaks down a sparse recovery problem into a series of subset linear least-squares subproblems [52, §6.4]. Here, we extend the concept of NLS to the single-look multi-master data model (12) and address the subproblem (15), or equivalently

$$\min_{\mathbf{x}} \frac{1}{2}\|(\mathbf{A}\mathbf{x}) \circ \left(\overline{\mathbf{B}\mathbf{x}}\right) - \mathbf{b}\|_2^2 \tag{17}$$

where $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{m \times n}$, $\mathbf{x} \in \mathbb{C}^n$, and $\mathbf{b} \in \mathbb{C}^m$ with $m > n$. As can be concluded from Section II-C, (17) is clearly of interest, since it not only solves the nonconvex sparse recovery problem (13) but also underlies the model-order selection (14).

A. Algorithm

In this section, we develop two algorithms for solving (17).

The first algorithm is based on the alternating direction method of multipliers (ADMM) [53]. The ADMM solves a minimization problem by alternatively minimizing its augmented Lagrangian [54, p. 509], in which the augmentation term is scaled by a penalty parameter $\rho \in \mathbb{R}_{++}$. A short recap can be found in Appendix A. It converges under very general conditions with medium accuracy [53, §3.2].

Now, we consider (17) in its equivalent form

$$\begin{aligned} \min_{\mathbf{x},\mathbf{z}} \ & \frac{1}{2}\|(\mathbf{A}\mathbf{x}) \circ \left(\overline{\mathbf{B}\mathbf{z}}\right) - \mathbf{b}\|_2^2 \\ \text{s.t.} \ & \mathbf{x} - \mathbf{z} = \mathbf{0}. \end{aligned} \tag{18}$$

This is essentially a bi-convex problem with an affine constraint [53, §9.2]. Applying the ADMM update rules leads to Algorithm 1. Note that both $\mathbf{x}$- and $\mathbf{z}$-updates boil down to solving linear least-squares problems.

Algorithm 1 ADMM-Based Algorithm for Solving (17)

Input: $\mathbf{A}$, $\mathbf{B}$, $\mathbf{b}$, $\mathbf{z}^{(0)}$, ρ
Initialize $\mathbf{z} \leftarrow \mathbf{z}^{(0)}$
Until stopping criterion is satisfied, **Do**
  $\tilde{\mathbf{A}} \leftarrow \operatorname{Diag}(\overline{\mathbf{B}\mathbf{z}})\mathbf{A}$
  $\mathbf{x} \leftarrow (\tilde{\mathbf{A}}^H\tilde{\mathbf{A}} + \rho\mathbf{I})^{-1}(\tilde{\mathbf{A}}^H\mathbf{b} + \rho\mathbf{z} - \mathbf{y})$
  $\tilde{\mathbf{B}} \leftarrow \operatorname{Diag}(\overline{\mathbf{A}\mathbf{x}})\mathbf{B}$
  $\mathbf{z} \leftarrow (\tilde{\mathbf{B}}^H\tilde{\mathbf{B}} + \rho\mathbf{I})^{-1}(\tilde{\mathbf{B}}^H\overline{\mathbf{b}} + \rho\mathbf{x} + \mathbf{y})$
  $\mathbf{y} \leftarrow \mathbf{y} + \rho(\mathbf{x} - \mathbf{z})$
Output: $\mathbf{z}$

The second algorithm uses trust-region Newton's method that exploits the second-order information for solving the general unconstrained nonlinear minimization problems [55, §4]. The rationale behind this choice is to circumvent the saddle points that cannot be identified by the first-order information [56]. In each iteration, a norm ball or "trust-region" centered at the current iterate is adaptively chosen.

If the second-order Taylor polynomial of the objective function is sufficiently good for approximation, a descent direction is found by solving a quadratically constrained quadratic minimization problem. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is the objective function, the subproblem at the iterate $\mathbf{x} \in \mathbb{R}^n$ is

$$\begin{aligned} &\min_{\Delta\mathbf{x}} \; f(\mathbf{x}) + \nabla f(\mathbf{x})^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \nabla^2 f(\mathbf{x}) \Delta\mathbf{x} \\ &\text{s.t.} \;\; \|\Delta\mathbf{x}\|_2 \le r \end{aligned} \tag{19}$$

where $\Delta\mathbf{x} \in \mathbb{R}^n$ is the search direction, ∇f and $\nabla^2 f$ denote the gradient and Hessian of f, respectively, and $r \in \mathbb{R}_{++}$ is the current trust-region radius. By means of the Karush–Kuhn–Tucker (KKT) conditions for the nonconvex problems, Nocedal and Wright [55, §4.3] divided (19) into several cases: in one case, a 1-D root-finding problem with respect to the dual variable is solved by using, for example, Newton's method, while in the others, the solutions are analytic. Since the technical details are quite overwhelming, we do not intend to provide an exposition here. Interested readers are advised to refer to [55, §4.3]. It can be shown that trust-region Newton's method converges to a critical point with high accuracy under general conditions [55, p. 92].

By verifying the Cauchy–Riemann equations (see [57, p. 50]), it is easy to show that the objective function of (17) is not complex-differentiable with respect to $\mathbf{x}$. In lieu of using Wirtinger differentiation that does not contain all the second-order information, we exploit the fact that the mapping $\mathbf{x} \mapsto (\mathbf{x}_R, \mathbf{x}_I)$ is isomorphic and let

$$f(\mathbf{x}_R, \mathbf{x}_I) \coloneqq \frac{1}{2}\|(\mathbf{A}\mathbf{x}) \circ \big(\overline{\mathbf{B}\mathbf{x}}\big) - \mathbf{b}\|_2^2 \tag{20}$$

where $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is real-differentiable with respect to $\mathbf{x}_R$ and $\mathbf{x}_I$. Straightforward computations reveal its gradient as

$$\nabla f(\mathbf{x}_R, \mathbf{x}_I) = \begin{pmatrix} \dfrac{\partial f}{\partial \mathbf{x}_R} \\ \dfrac{\partial f}{\partial \mathbf{x}_I} \end{pmatrix} = \begin{pmatrix} \Re(\mathbf{d}) \\ \Im(\mathbf{d}) \end{pmatrix} \tag{21}$$

where

$$\begin{aligned} \mathbf{d} \coloneqq \; & \mathbf{A}^H \Big(\big((\mathbf{A}\mathbf{x}) \circ \big(\overline{\mathbf{B}\mathbf{x}}\big) - \mathbf{b} \big) \circ (\mathbf{B}\mathbf{x}) \Big) \\ & + \mathbf{B}^H \Big(\big(\big(\overline{\mathbf{A}\mathbf{x}}\big) \circ (\mathbf{B}\mathbf{x}) - \overline{\mathbf{b}} \big) \circ (\mathbf{A}\mathbf{x}) \Big) \end{aligned} \tag{22}$$

and its Hessian as

$$\begin{aligned} \nabla^2 f(\mathbf{x}_R, \mathbf{x}_I) &= \begin{pmatrix} \dfrac{\partial^2 f}{\partial \mathbf{x}_R^2} & \dfrac{\partial^2 f}{\partial \mathbf{x}_R \partial \mathbf{x}_I} \\ \dfrac{\partial^2 f}{\partial \mathbf{x}_I \partial \mathbf{x}_R} & \dfrac{\partial^2 f}{\partial \mathbf{x}_I^2} \end{pmatrix} \\ &= \begin{pmatrix} \Re(\mathbf{C}+\mathbf{D}+\mathbf{E}) & -\Im(\mathbf{C}-\mathbf{D}+\mathbf{E}) \\ \Im(\mathbf{C}+\mathbf{D}+\mathbf{E}) & \Re(\mathbf{C}-\mathbf{D}+\mathbf{E}) \end{pmatrix} \end{aligned} \tag{23}$$

where

$$\begin{aligned} \mathbf{C} &\coloneqq \mathbf{A}^H \operatorname{Diag}\big((\mathbf{B}\mathbf{x}) \circ \big(\overline{\mathbf{B}\mathbf{x}}\big)\big)\mathbf{A} + \mathbf{B}^H \operatorname{Diag}\big((\mathbf{A}\mathbf{x}) \circ \big(\overline{\mathbf{A}\mathbf{x}}\big)\big)\mathbf{B} \\ \mathbf{D} &\coloneqq \mathbf{A}^H \operatorname{Diag}((\mathbf{A}\mathbf{x}) \circ (\mathbf{B}\mathbf{x}))\overline{\mathbf{B}} + \mathbf{B}^H \operatorname{Diag}((\mathbf{A}\mathbf{x}) \circ (\mathbf{B}\mathbf{x}))\overline{\mathbf{A}} \\ \mathbf{E} &\coloneqq \mathbf{A}^H \operatorname{Diag}\big((\mathbf{A}\mathbf{x}) \circ \big(\overline{\mathbf{B}\mathbf{x}}\big) - \mathbf{b}\big)\mathbf{B} \\ &\quad + \mathbf{B}^H \operatorname{Diag}\big(\big(\overline{\mathbf{A}\mathbf{x}}\big) \circ (\mathbf{B}\mathbf{x}) - \overline{\mathbf{b}}\big)\mathbf{A}. \end{aligned} \tag{24}$$

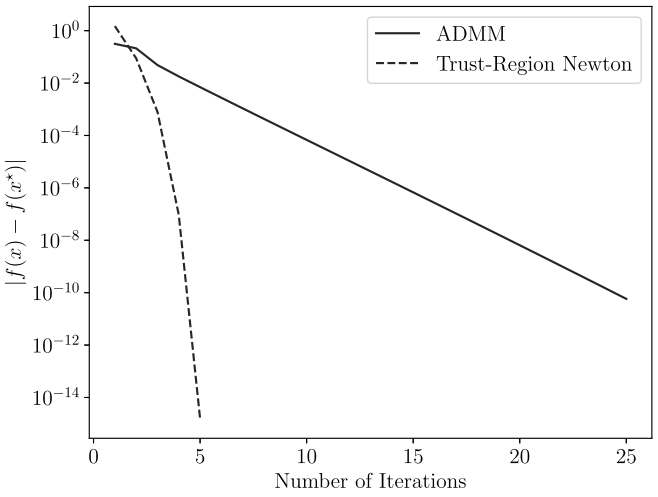


Fig. 2. Convergence curve of NLS using the ADMM (solid line) and trust-region Newton's method (dashed line).

Note that $\mathbf{d} : \mathbb{C}^n \to \mathbb{C}^n$ and $\mathbf{C}, \mathbf{D}, \mathbf{E} : \mathbb{C}^n \to \mathbb{C}^{n\times n}$ are essentially the functions of $\mathbf{x}$. Here, we drop the parentheses in order to simplify the notation. For the same purpose, we adopt the following convention:

$$f(\mathbf{x}) \coloneqq f(\Re(\mathbf{x}), \Im(\mathbf{x})) = f(\mathbf{x}_R, \mathbf{x}_I). \tag{25}$$

By using the first- and the second-order information of (20), (17) can be directly tackled by trust-region Newton's method by solving a sequence of subproblems in the form of (19). For any optimal point $\mathbf{x}^\star$, the KKT condition is

$$\nabla f\big(\mathbf{x}^\star\big) = \mathbf{0} \iff \mathbf{d}\big(\mathbf{x}^\star\big) = \mathbf{0}. \tag{26}$$

Remark: Likewise, the objective function of (16) is real-differentiable with respect to ${\gamma_l}^R$, ${\gamma_l}^I$ and s_l, $\forall l \in [\hat{K}]$. Therefore, trust-region Newton's method is directly applicable. Alternatively, the first-order methods such as Broyden–Fletcher–Goldfarb–Shanno (BFGS, see [55, §6.1] and the references therein) can also be used.

Needless to say, it is not guaranteed that these algorithms always converge to a global minimum. We will demonstrate later in Section V that the solutions are often good enough. Fig. 2 shows the typical convergence curves in the case of double scatterers (#6 in Section V-C). In order to generate this plot, we first let one algorithm run nonstop until it converged with very high precision. We then took this solution as an optimal point $\mathbf{x}^\star$ and compared the absolute difference of the objective value $|f(\mathbf{x}) - f(\mathbf{x}^\star)|$. Both ADMM and trust-region Newton's method converged to the same solution (up to a constant phase angle; see Section III-B), although it only took the latter less than ten iterations. Still, the former can be interesting due to the simplicity of its update rules (see Algorithm 1). In Section V, the latter will be used for demonstration purposes.

B. Analysis of the Objective Function

Due to the nonconvexity of the objective function (20), its analysis is not straightforward. We are primarily concerned with the following two questions.

1) Under which circumstances do critical points or local extrema exist?
2) If they do exist, how many are they?

This section shall provide a partial answer to these questions.

First, we state the following general observation.

Proposition 1: For any $\mathbf{x} \in \mathbb{C}^n$ and $\phi \in \mathbb{R}$, any eigenvalue of $\nabla^2 f(\mathbf{x})$ is also an eigenvalue of $\nabla^2 f(\mathbf{x}\exp(j\phi))$ and vice versa.

Proof: See Appendix B. □

Informally, this proposition implies that the definiteness of the Hessian is invariant under any rotation with a constant phase angle.

Now, we state the main theorem for the general case.

Theorem 2: The properties of the critical points of $f(\mathbf{x})$.

1) $\mathbf{0}$ is a critical point: it is a local minimum if $\mathbf{A}^H \operatorname{Diag}(\mathbf{b})\mathbf{B} + \mathbf{B}^H \operatorname{Diag}(\overline{\mathbf{b}})\mathbf{A} \prec \mathbf{0}$ and a local maximum if $\mathbf{A}^H \operatorname{Diag}(\mathbf{b})\mathbf{B} + \mathbf{B}^H \operatorname{Diag}(\overline{\mathbf{b}})\mathbf{A} \succ \mathbf{0}$.
2) If there exists a *nonzero* critical point, then $\mathbf{A}^H \operatorname{Diag}(\mathbf{b})\mathbf{B} + \mathbf{B}^H \operatorname{Diag}(\overline{\mathbf{b}})\mathbf{A} \not\prec \mathbf{0}$.
3) Suppose there exists a *nonzero* critical point $\mathbf{z}$, then the following holds.
   a) $\nabla^2 f(\mathbf{z})$ is rank-deficient.
   b) There exist an infinite number of critical points in the form of $\mathbf{z}\exp(j\phi)$, $\phi \in \mathbb{R} \setminus \{0\}$. Each has the same objective function value as $\mathbf{z}$, and its Hessian has the same definiteness.

Proof: See Appendix C. □

This theorem implies that if there exists one critical point, then there are an infinite number of them up to a constant phase angle, and each is good. Furthermore, we conjecture that $\mathbf{A}^H \operatorname{Diag}(\mathbf{b})\mathbf{B} + \mathbf{B}^H \operatorname{Diag}(\overline{\mathbf{b}})\mathbf{A} \succ \mathbf{0}$ is a necessary and sufficient condition [see Theorem 2(2)], and each *nonzero* critical point is also a local minimum under some mild conditions.

For the special case $n = 1$, i.e., $\mathbf{A}, \mathbf{B} \in \mathbb{C}^m$, $\mathbf{x} \in \mathbb{C}$, we have a much stronger result.

Theorem 3 ($n = 1$): The properties of the critical points of $f(\mathbf{x})$.

1) $\mathbf{0}$ is a critical point: it is a local minimum if $\Re((\mathbf{A} \circ \overline{\mathbf{B}})^H \mathbf{b}) < 0$ and a local maximum if $\Re((\mathbf{A} \circ \overline{\mathbf{B}})^H \mathbf{b}) > 0$.
2) There exists a *nonzero* critical point if and only if $\Re((\mathbf{A} \circ \overline{\mathbf{B}})^H \mathbf{b}) > 0$.
3) Suppose there exists a *nonzero* critical point $\mathbf{z}$, then the following holds.
   a) $\nabla^2 f(\mathbf{z})$ is positive-semidefinite and rank-deficient.[2]
   b) There exist an infinite number of critical points in the form of $\mathbf{z}\exp(j\phi)$, $\phi \in \mathbb{R} \setminus \{0\}$. Each has the same objective function value as $\mathbf{z}$, and its Hessian has the same definiteness.
   c) $\mathbf{z}$ is a local minimum.

Proof: See Appendix D. □

As a result, a *nonzero local minimum* exists if and only if $\Re((\mathbf{A} \circ \overline{\mathbf{B}})^H \mathbf{b}) > 0$. If this condition is satisfied, then there are infinitely many local minima that are exactly as good. Fig. 3 shows as an example the negative logarithm of (20) with a circle of local maxima.

[2]Note that $\nabla^2 f(\mathbf{z}) \in \mathbb{R}^{2\times 2}$ by definition.

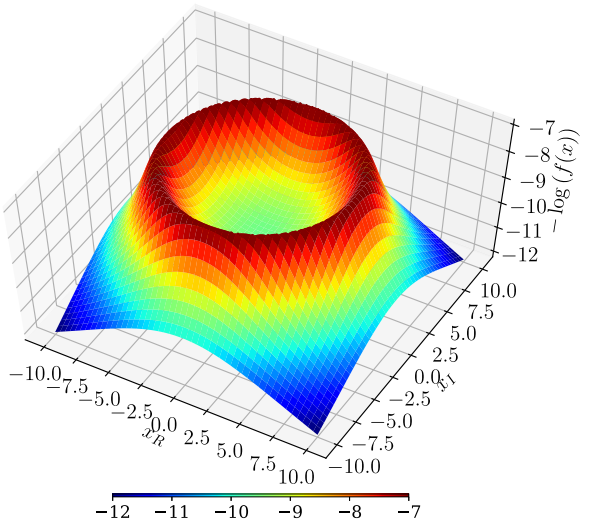


Fig. 3. Negative logarithm of the NLS objective function ($n = 1$) with a circle of local maxima at the verge of the "crater."

Finally, Theorem 3 implies the following interesting result.

Corollary 4 ($n = 1$): Each *nonzero local minimum* (if it exists) is given by

$$\mathbf{z} = \frac{\Re\left(\left(\mathbf{A} \circ \overline{\mathbf{B}}\right)^H \mathbf{b}\right)^{1/2}}{\|\mathbf{A} \circ \overline{\mathbf{B}}\|_2} \exp(j\phi) \tag{27}$$

for some $\phi \in \mathbb{R}$.

Proof: See the proof of Theorem 3(2). □

Now we return to our problem in SAR tomography. For the single-look multi-master data model (12), this corollary motivates the 1-D spectral estimator

$$|\hat{\gamma}_l| := \begin{cases} \dfrac{\Re\left((\mathbf{r}_l \circ \overline{\mathbf{s}_l})^H \mathbf{g}\right)^{1/2}}{\|\mathbf{r}_l \circ \overline{\mathbf{s}_l}\|_2}, & \text{if } \Re\left((\mathbf{r}_l \circ \overline{\mathbf{s}_l})^H \mathbf{g}\right) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{28}$$

$\forall l \in [L]$. Note that this also provides the solution for any 1-D NLS subproblem up to a constant phase angle. In the case of multiple scatterers, this estimator does not have any super-resolution power.

IV. Bi-Convex Relaxation and Alternating Minimization

This section introduces a second algorithm for solving the nonconvex sparse-recovery problem (13).

A. Algorithm

As a starting point, we replace the constraint in (13) with a sparsity-inducing regularization term, for example

$$\min_{\boldsymbol{\gamma}} \frac{1}{2}\|(\mathbf{R}\boldsymbol{\gamma}) \circ \left(\overline{\mathbf{S}\boldsymbol{\gamma}}\right) - \mathbf{g}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1 \tag{29}$$

where $\lambda \in \mathbb{R}_{++}$ trades the model goodness of fit for sparsity. In light of (26), the necessary condition for being an optimal

point $\boldsymbol{\gamma}^\star$ is

$$\lambda\partial\|\boldsymbol{\gamma}^\star\|_1 \ni \mathbf{R}^H\big((\mathbf{g}-(\mathbf{R}\boldsymbol{\gamma}^\star)\circ(\overline{\mathbf{S}\boldsymbol{\gamma}^\star})\big)\circ(\mathbf{S}\boldsymbol{\gamma}^\star)\big) + \mathbf{S}^H\big((\overline{\mathbf{g}}-(\overline{\mathbf{R}\boldsymbol{\gamma}^\star})\circ(\mathbf{S}\boldsymbol{\gamma}^\star)\big)\circ(\mathbf{R}\boldsymbol{\gamma}^\star)\big) \quad (30)$$

that is, the right-hand side is a subgradient of the ℓ_1 norm at $\boldsymbol{\gamma}^\star$. Obviously, $\mathbf{0}$ always satisfies this condition.

In principle, an ADMM-based algorithm similar to Algorithm 1 can be used to solve (29). However, our experience with the real SAR tomographic data shows that it often diverges, presumably due to the high mutual coherence of $\mathbf{R}$ and $\mathbf{S}$ under nonconvexity. For this reason, we consider instead the following relaxed version of (29):

$$\min_{\boldsymbol{\gamma},\boldsymbol{\theta}} \ \frac{1}{2}\|(\mathbf{R}\boldsymbol{\gamma})\circ(\overline{\mathbf{S}\boldsymbol{\theta}})-\mathbf{g}\|_2^2+\frac{\lambda_1}{2}\|\boldsymbol{\gamma}-\boldsymbol{\theta}\|_2^2+\lambda_2\|\begin{pmatrix}\boldsymbol{\gamma} & \boldsymbol{\theta}\end{pmatrix}\|_{1,2} \quad (31)$$

where $\lambda_1, \lambda_2 \in \mathbb{R}_{++}$. The objective function $\mathbb{C}^L \times \mathbb{C}^L \to \mathbb{R}$ is bi-convex, i.e., it is convex in $\boldsymbol{\gamma}$ with $\boldsymbol{\theta}$ fixed and convex in $\boldsymbol{\theta}$ with $\boldsymbol{\gamma}$ fixed. The first regularization term enforces $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ to have similar entries, and the second one promotes the same support. Since (31) is essentially an unconstrained bi-convex problem, it can be solved by using alternating minimization by Algorithm 2 (see also [58]–[60]).

Algorithm 2 Alternating Algorithm for Solving (31)

1: **Input:** $\mathbf{R}$, $\mathbf{S}$, $\mathbf{g}$, $\boldsymbol{\gamma}^{(0)}$, λ_1, λ_2
2: **Initialize** $\boldsymbol{\gamma} \leftarrow \boldsymbol{\gamma}^{(0)}$
3: **Until** stopping criterion is satisfied, **Do**
4: $\quad\tilde{\mathbf{S}} \leftarrow \mathrm{Diag}(\overline{\mathbf{R}\boldsymbol{\gamma}})\mathbf{S}$
5: $\quad\boldsymbol{\theta} \leftarrow \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\|\tilde{\mathbf{S}}\boldsymbol{\theta}-\overline{\mathbf{g}}\|_2^2+\frac{\lambda_1}{2}\|\boldsymbol{\theta}-\boldsymbol{\gamma}\|_2^2+\lambda_2\|\begin{pmatrix}\boldsymbol{\theta} & \boldsymbol{\gamma}\end{pmatrix}\|_{1,2}$
6: $\quad\tilde{\mathbf{R}} \leftarrow \mathrm{Diag}(\overline{\mathbf{S}\boldsymbol{\theta}})\mathbf{R}$
7: $\quad\boldsymbol{\gamma} \leftarrow \arg\min_{\boldsymbol{\gamma}} \frac{1}{2}\|\tilde{\mathbf{R}}\boldsymbol{\gamma}-\mathbf{g}\|_2^2+\frac{\lambda_1}{2}\|\boldsymbol{\gamma}-\boldsymbol{\theta}\|_2^2+\lambda_2\|\begin{pmatrix}\boldsymbol{\gamma} & \boldsymbol{\theta}\end{pmatrix}\|_{1,2}$
8: **Output:** $\boldsymbol{\gamma}$

Each time when either $\boldsymbol{\gamma}$ or $\boldsymbol{\theta}$ is fixed, it becomes a convex problem in the generic form of

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2+\frac{\lambda_1}{2}\|\mathbf{x}-\mathbf{u}\|_2^2+\lambda_2\|\begin{pmatrix}\mathbf{x} & \mathbf{u}\end{pmatrix}\|_{1,2} \quad (32)$$

or equivalently

$$\min_{\mathbf{x},\mathbf{Z}} \ \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2+\frac{\lambda_1}{2}\|\mathbf{x}-\mathbf{u}\|_2^2+\lambda_2\|\mathbf{Z}\|_{1,2} \quad \text{s.t.} \ \begin{pmatrix}\mathbf{x} & \mathbf{u}\end{pmatrix}-\mathbf{Z}=\mathbf{0} \quad (33)$$

where $\mathbf{Z} \in \mathbb{C}^{L\times 2}$. Applying the ADMM update rules leads to Algorithm 3.

Algorithm 3 ADMM-Based Algorithm for Solving (32)

1: **Input:** $\mathbf{A}$, $\mathbf{b}$, $\mathbf{u}$, $\mathbf{Z}^{(0)}$, λ_1, λ_2, ρ
2: **Initialize** $\mathbf{Z} \leftarrow \mathbf{Z}^{(0)}$
3: **Until** stopping criterion is satisfied, **Do**
4: $\quad\mathbf{x} \leftarrow (\mathbf{A}^H\mathbf{A}+(\lambda_1+\rho)\mathbf{I})^{-1}(\mathbf{A}^H\mathbf{b}+\lambda_1\mathbf{u}+\rho\mathbf{z}_1-\mathbf{y}_1)$
5: $\quad\mathbf{Z} \leftarrow \mathrm{Prox}_{\ell_{1,2},\lambda_2/\rho}\big(\begin{pmatrix}\mathbf{x} & \mathbf{u}\end{pmatrix}+(1/\rho)\mathbf{Y}\big)$
6: $\quad\mathbf{Y} \leftarrow \mathbf{Y}+\rho\big(\begin{pmatrix}\mathbf{x} & \mathbf{u}\end{pmatrix}-\mathbf{Z}\big)$
7: **Output:** $\mathbf{z}_1$

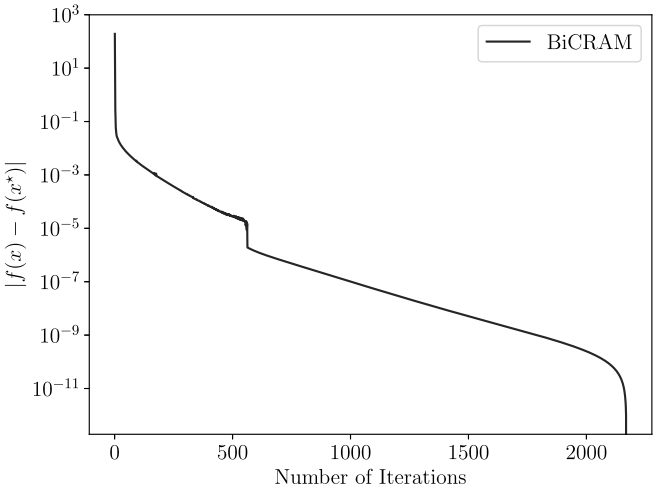


Fig. 4. Convergence curve of BiCRAM. The horizontal axis refers to the outer iterations in Algorithm 2.

$\mathrm{Prox}_{\ell_{1,2},\lambda} : \mathbb{C}^{L\times 2} \to \mathbb{C}^{L\times 2}$ is the proximal operator of the $\ell_{1,2}$ norm scaled by λ [61], that is

$$\mathrm{Prox}_{\ell_{1,2},\lambda}(\mathbf{X}) \coloneqq \arg\min_{\mathbf{Z}} \ \lambda\|\mathbf{Z}\|_{1,2}+\frac{1}{2}\|\mathbf{X}-\mathbf{Z}\|_F^2 \quad (34)$$

whose ith row is given by [61, §6.5.4]

$$\mathrm{Prox}_{\ell_{1,2},\lambda}(\mathbf{X})^i = \big(1-\lambda/\|\mathbf{x}^i\|_2\big)_+\mathbf{x}^i \quad (35)$$

where $(x)_+ \coloneqq \max(x, 0)$. This proximal operator promotes (the columns of) $\mathbf{Z}$ to be jointly sparse and, therefore, $\mathbf{x}$ to share the same support with $\mathbf{u}$.

Due to the nonconvexity of (31), it is very difficult to establish a convergence guarantee for Algorithm 2 from a theoretical point of view. However, our experiments with the real SAR tomographic data (see Section V) show that it converges empirically. As an example, Fig. 4 depicts a convergence curve in the case of two scatterers that are closely located (#6 in Section V-C).

In terms of regularization parameter tuning, we adopt the approach of sampling the solution path $(\lambda_1, \lambda_2) \mapsto \mathbf{x}$ and selecting the solution with the highest penalized likelihood (14). Last, this procedure can be simplified by performing 1-D search, i.e., fixing one parameter and tuning the other at a time.

B. Implementation

This section addresses several implementation aspects that contribute to accelerating Algorithm 3 (and therefore Algorithm 2). The exposition is based on an ADMM-based algorithm for solving the ℓ_1-regularized least-squares (L1RLS) problem

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2+\lambda\|\mathbf{x}\|_1. \quad (36)$$

This is more suitable for demonstrating the power of different acceleration techniques, since each of its subproblems has an analytical solution and does not involve iteratively solving another optimization problem (see Algorithm 2). In addition, it will also be used as a reference in Section V.

Now, consider (36) in its equivalent form

$$\min_{\mathbf{x},\mathbf{z}} \ \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2+\lambda\|\mathbf{z}\|_1 \quad \text{s.t.} \ \ \mathbf{x}-\mathbf{z}=\mathbf{0}. \tag{37}$$

Applying the ADMM update rules leads to Algorithm 4.

Algorithm 4 ADMM-Based Algorithm for Solving (36)

Input: $\mathbf{A}$, $\mathbf{b}$, $\mathbf{z}^{(0)}$, λ, ρ
Initialize $\mathbf{z} \leftarrow \mathbf{z}^{(0)}$
Until stopping criterion is satisfied, **Do**
  $\mathbf{x} \leftarrow (\mathbf{A}^H\mathbf{A}+\rho\mathbf{I})^{-1}(\mathbf{A}^H\mathbf{b}+\rho\mathbf{z}-\mathbf{y})$
  $\mathbf{z} \leftarrow \mathrm{Prox}_{\ell_1,\lambda/\rho}(\mathbf{x}+(1/\rho)\mathbf{y})$
  $\mathbf{y} \leftarrow \mathbf{y}+\rho(\mathbf{x}-\mathbf{z})$
Output: $\mathbf{z}$

Likewise, $\mathrm{Prox}_{\ell_1,\lambda}:\mathbb{C}^L\to\mathbb{C}^L$ is the proximal operator of the ℓ_1 norm scaled by λ (also known as the soft-thresholding operator [62])

$$\mathrm{Prox}_{\ell_1,\lambda}(\mathbf{x}) \coloneqq \arg\min_{\mathbf{z}} \ \lambda\|\mathbf{z}\|_1+\frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2 \tag{38}$$

whose ith entry is given by [61, §6.5.2]

$$\mathrm{Prox}_{\ell_1,\lambda}(\mathbf{x})_i = (1-\lambda/|x_i|)_+ x_i. \tag{39}$$

The first technique provides an easier way for the $\mathbf{x}$-update.

1) Matrix-Inversion Lemma: In Algorithms 3 and 4, an L-by-L matrix needs to be inverted. For large L, a direct exact approach can be tedious. Instead, we exploit the following lemma.

Lemma 5 (Matrix Inversion Lemma [63]): For any $\mathbf{A}\in\mathbb{C}^{n\times m}$, $\mathbf{B}\in\mathbb{C}^{m\times n}$ and nonsingular $\mathbf{C}\in\mathbb{C}^{n\times n}$, we have

$$(\mathbf{A}\mathbf{B}+\mathbf{C})^{-1}=\mathbf{C}^{-1}-\mathbf{C}^{-1}\mathbf{A}\big(\mathbf{I}+\mathbf{B}\mathbf{C}^{-1}\mathbf{A}\big)^{-1}\mathbf{B}\mathbf{C}^{-1}. \tag{40}$$

Lemma 5 suggests a more efficient method if inverting $\mathbf{C}$ is straightforward. This is the case for the matrices in the form of $\mathbf{A}^H\mathbf{A}+\rho\mathbf{I}$ since

$$\big(\mathbf{A}^H\mathbf{A}+\rho\mathbf{I}\big)^{-1}=\frac{1}{\rho}\mathbf{I}-\frac{1}{\rho^2}\mathbf{A}^H\left(\mathbf{I}+\frac{1}{\rho}\mathbf{A}\mathbf{A}^H\right)^{-1}\mathbf{A} \tag{41}$$

that is, instead of the original L-by-L matrix, only an N'-by-N' matrix needs to be inverted.

Alternatively, the linear least-squares (sub)problems can be solved iteratively in order to deliver an approximate solution [64], which is known as inexact minimization [53, §3.4.4].

The following techniques can be employed to improve convergence.

2) Varying Penalty Parameter: The penalty parameter ρ can be updated at each iteration. Besides the convergence aspect, this also renders Algorithm 4 to be less dependent on the initial choice of ρ. A common heuristic [53, §3.4.1] is to set

$$\rho^{(k+1)} \coloneqq \begin{cases} \tau\rho^{(k)}, & \text{if } \|\mathbf{r}^{(k)}\|_2 > \mu\|\mathbf{s}^{(k)}\|_2 \\ \rho^{(k)}/\tau, & \text{if } \|\mathbf{s}^{(k)}\|_2 > \mu\|\mathbf{r}^{(k)}\|_2 \\ \rho^{(k)}, & \text{otherwise} \end{cases} \tag{42}$$

at the $(k+1)$th iteration, where $\tau,\mu>1$ are the parameters, $\mathbf{r}^{(k)}\coloneqq\mathbf{x}^{(k)}-\mathbf{z}^{(k)}$ is the primal residual, and $\mathbf{s}^{(k)}\coloneqq\rho^{(k)}(\mathbf{z}^{(k)}-\mathbf{z}^{(k-1)})$ is the dual residual. As $k\to\infty$, $\mathbf{r}^{(k)}$ and $\mathbf{s}^{(k)}$ both converge to $\mathbf{0}$. Intuitively, increasing ρ tends to put a larger penalty on the augmenting term $(\rho/2)\|\mathbf{x}-\mathbf{z}\|_2^2$ in the augmented Lagrangian and, consequently, decreases $\|\mathbf{r}^{(k)}\|_2$ on the one hand, and to increase $\|\mathbf{s}^{(k)}\|_2$ by definition on the other and vice versa. The rationale is to balance $\mathbf{r}^{(k)}$ and $\mathbf{s}^{(k)}$ so that they are approximately of the same order. Naturally, one downside is that (41) needs to be recomputed whenever ρ changes.

3) Diagonal Preconditioning: The augmenting term $(\rho/2)\|\mathbf{x}-\mathbf{z}\|_2^2$ in the augmented Lagrangian can be replaced by

$$(1/2)\langle\mathbf{P}(\mathbf{x}-\mathbf{z}),\mathbf{x}-\mathbf{z}\rangle \tag{43}$$

where $\mathbf{P}\succ\mathbf{0}$ is a real diagonal matrix. Note that this falls under the category of more general augmenting terms [53, §3.4.2]. By means of this, Algorithm 4 is deprived of the burden of choosing ρ and the ADMM updates become

$$\begin{aligned} \mathbf{x} &\leftarrow \big(\mathbf{A}^H\mathbf{A}+\mathbf{P}\big)^{-1}\big(\mathbf{A}^H\mathbf{b}+\mathbf{P}\mathbf{z}-\mathbf{y}\big) \\ \mathbf{z} &\leftarrow \mathrm{Prox}_{\ell_1,\lambda/\mathbf{p}}\big(\mathbf{x}+\mathbf{P}^{-1}\mathbf{y}\big) \\ \mathbf{y} &\leftarrow \mathbf{y}+\mathbf{P}(\mathbf{x}-\mathbf{z}) \end{aligned} \tag{44}$$

where $\mathbf{p}\coloneqq\mathrm{Diag}(\mathbf{P})$, and $\mathrm{Prox}_{\ell_1,\mathbf{w}}:\mathbb{C}^L\to\mathbb{C}^L$ is the proximal operator of the weighted ℓ_1 norm with weights $\mathbf{w}\in\mathbb{R}^L_{++}$

$$\mathrm{Prox}_{\ell_1,\mathbf{w}}(\mathbf{x}) \coloneqq \arg\min_{\mathbf{z}} \ \|\mathbf{z}\|_{\mathbf{w},1}+\frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2 \tag{45}$$

whose ith entry is given by

$$\mathrm{Prox}_{\ell_1,\mathbf{w}}(\mathbf{x})_i = (1-w_i/|x_i|)_+ x_i. \tag{46}$$

In case $\mathbf{A}^H\mathbf{A}$ is ill-conditioned (such as in SAR tomography), $\mathbf{P}$ can be interpreted as a preconditioner. Needless to say, Lemma 5 can also be applied to invert $\mathbf{A}^H\mathbf{A}+\mathbf{P}$.

Pock and Chambolle (2011) [65] proposed a simple and elegant way to construct diagonal preconditioners for a primal–dual algorithm [54, §15.2] with guaranteed convergence

$$p_i \coloneqq 1/\|\mathbf{a}_i\|_\alpha^\alpha \quad \forall i\in[L] \tag{47}$$

where $\alpha\in[0,2]$ is a parameter.

4) Over-Relaxation: This means inserting between the $\mathbf{x}$- and $\mathbf{z}$-updates of Algorithm 4 the following additional update:

$$\mathbf{x} \leftarrow \beta\mathbf{x}+(1-\beta)\mathbf{z} \tag{48}$$

where $\beta\in[1.5,1.8]$ (see [53, §3.4.3] and the references therein).

Fig. 5 shows the convergence curve of Algorithm 4 using different acceleration techniques, as applied to the real SAR tomographic data (#6 in Section V). Each technique did contribute to accelerating Algorithm 4 in comparison with "baseline," where we set $\rho=1$. The number of iterations of this and five other cases is listed in Table II. Obviously, the combination of diagonal preconditioning and over-relaxation was the most competitive one and will, therefore, be adopted for all the ADMM-based algorithms in the following.

TABLE II
NUMBER OF ITERATIONS USING DIFFERENT ACCELERATION TECHNIQUES

	#1	#2	#3	#4	#5	#6
Baseline	20289	14250	15752	19571	23920	16983
Vary	1445	1101	1780	1958	2005	1033
Precondition	679	480	504	1074	840	805
Relax	11272	7916	8750	10866	13288	9446
Vary+Relax	767	625	871	821	834	1051
Precondition+Relax	377	260	288	595	467	447

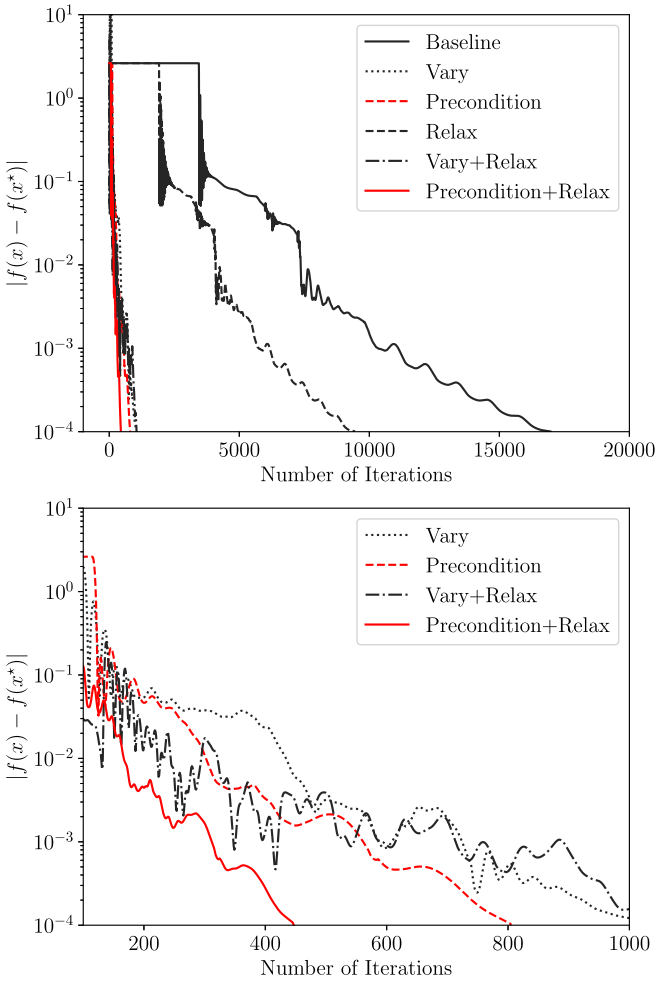


Fig. 5. Convergence curve of Algorithm 4 using different acceleration techniques. "Baseline": $\rho = 1$. "Vary": varying penalty parameter. "Precondition": diagonal preconditioning. "Relax": over-relaxation. (Bottom) Close-up of the top figure.

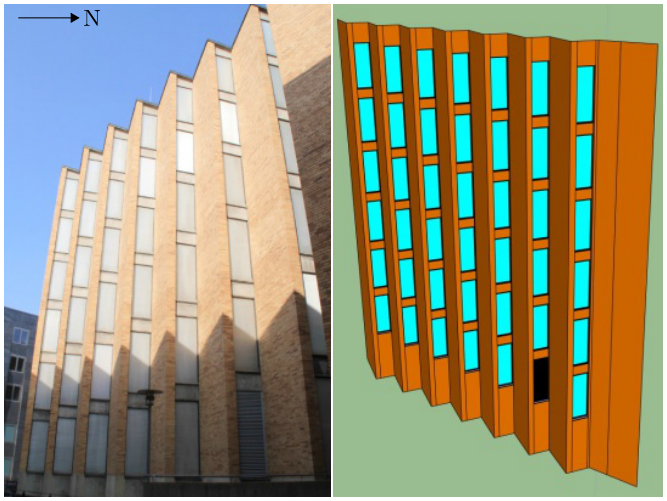


Fig. 6. Eastern facade of the six-storey TUM-Nordbau building in our region of interest [68]. (Left) *In situ* photograph. (Right) 3-D facade model. The black shape corresponds to a metallic window.

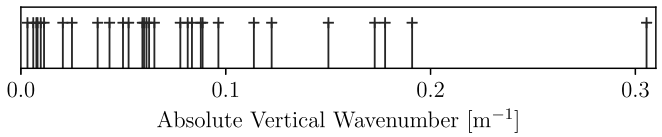


Fig. 7. Single-master absolute vertical wavenumbers. The largest one is approximately 0.31 m^{-1}.

V. EXPERIMENT WITH TERRASAR-X DATA

In this section, we report our experimental results with a real SAR data set.

A. Design of Experiment

As a demonstration, we used 31 TerraSAR-X staring spotlight repeat-pass acquisitions of the central Munich area from March 31, 2016 to December 7, 2017. This data set was processed with the DLR's Integrated Wide Area Processor [66], [67], as was elaborately described in [17]. In addition to sidelobe detection (see [17] and the references therein), any nonpeak point inside a main lobe was also removed, since it would otherwise lead to a "ghost" scatterer in the result, as any sidelobe point would do too. Our region of interest contains a six-storey building ("Nordbau") of the Technical University of Munich (TUM) shown in Fig. 6 (left). The building signature in the SAR intensity image can be observed in Fig. 9, where the regular grid of salient points within the building footprint is a result of triple reflections on three orthogonal surfaces: metal plate (behind window glass), window ledge, and brick wall [68]. After main and sidelobe detection, a total of 594 looks were left, whose azimuth-range positions are shown in Fig. 8 (bottom).

A *single-master* stack was formed by choosing the acquisition from December 20, 2016 as the one and only master. Its vertical wavenumbers are shown in Fig. 7. A sinusoidal basis function was used for modeling periodical motion induced by temperature change. The vertical Rayleigh resolution at the scene center is approximately 12.66 m. The Crámer–Rao lower bound (CRLB) of height estimates

given the aforementioned periodical deformation model [25] and a nominal signal-to-noise ratio (SNR) of 2 dB is approximately 1.10 m. NLS and L1RLS were applied to this stack for tomographic reconstruction. The latter was solved by Algorithm 4 augmented with diagonal preconditioning and over-relaxation (see Section IV-B), where we set $\beta = 1.8$ and the choice of α is irrelevant (since $\mathbf{A}$ is a Fourier matrix). The solution path of L1RLS was sampled 11 times with the regularization parameter varying logarithmically from $\lambda_{\min} \coloneqq 5 \cdot 10^{-2} \|\mathbf{R}^H \mathbf{g}\|_\infty$ to $\lambda_{\max} \coloneqq 5 \cdot 10^{-1} \|\mathbf{R}^H \mathbf{g}\|_\infty$.

We constructed a *multi-master* stack of small temporal baselines: suppose $1', 2', 3', 4', \ldots$ is a chronologically ordered sequence of SLCs, the interferograms (edges) are $(1', 2')$, $(3', 4')$, and so on (see Fig. 1). As a result, this stack consists of 15 interferograms. Due to the small-baseline feature of this stack, we did not employ any deformation model for the sake of simplicity. NLS and BiCRAM were applied to reconstruct the elevation profile, where the latter was solved by Algorithm 3, employing diagonal preconditioning and over-relaxation. Likewise, the solution path of BiCRAM was also sampled 11 times, where λ_1 was fixed as 1 (since it was deemed relatively insignificant as far as our experience went), and λ_2 was set to vary logarithmically from $\lambda_{\min} \coloneqq 5 \cdot 10^{-2} \max\{\|\mathbf{R}^H \mathbf{g}\|_\infty, \|\mathbf{S}^H \mathbf{g}\|_\infty\}$ to $\lambda_{\max} \coloneqq 5 \cdot 10^{-1} \max\{\|\mathbf{R}^H \mathbf{g}\|_\infty, \|\mathbf{S}^H \mathbf{g}\|_\infty\}$. The initial solution was given by $\boldsymbol{\gamma}^{(0)} = (\mathbf{R} \circ \overline{\mathbf{S}})^H \mathbf{g}$ due to its simplicity. Alternatively, (28) could be used. In terms of off-grid correction, forward-mode automatic differentiation [69] was employed in order to circumvent analytically differentiating the objective function of (16) for any number of scatterers, and the optimization problem was solved by means of a BFGS implementation [70].

Finally, we built a second small-baseline *multi-master* stack in the identical way as the previous one. In addition, we normalized each interferogram with the corresponding master amplitude. We will refer to this as the *fake single-master* stack, since we treated it as if it had been a *single-master* one. In order to apply the *single-master* approach, we calculated for each interferogram the difference between the slave and master wavenumbers and used it as if it had been the wavenumber baseline, i.e., by inadequately assuming

$$g_n \overline{g_m} \approx \sum_l \gamma_l \exp(-j(k_n - k_m) s_l) \tag{49}$$

for each $(m, n) \in \mathbf{E}(\mathbf{G})$. Needless to say, NLS and L1RLS were employed exactly the same as that in the single-master case.

The next section briefly explains how we generated the ground-truth data.

B. Generation of Height Ground Truth

Height ground-truth data were made available by an SAR imaging geodesy and simulation framework [68]. The starting point was to create a 3-D facade model from terrestrial measurements, by a (drone-borne) camera, tachymeter, measuring rod, and differential global positioning system (GPS), with an overall accuracy better than 2 cm and a very high level of details [68]. Ground control points were used for referencing this facade model to an international terrestrial reference frame. A visualization of the 3-D facade model is provided in Fig. 6 (right). The ray-tracing-based

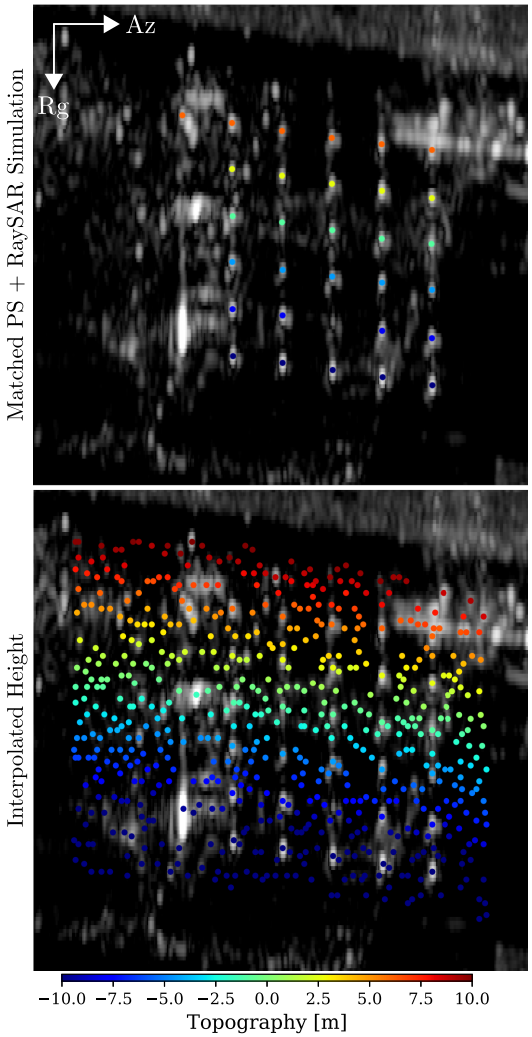


Fig. 8. SAR intensity image and height ground truth of our region of interest. (Top) RaySAR height simulations at 30 matched PS coordinates (Level 1). (Bottom) Interpolated height at 594 facade looks (Level 2).

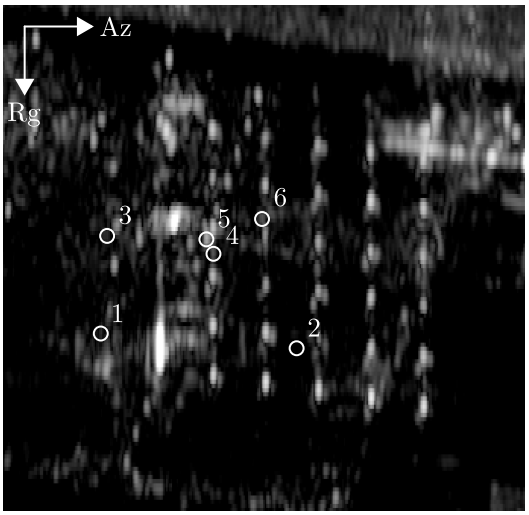


Fig. 9. Locations of six looks subject to facade-roof layover.

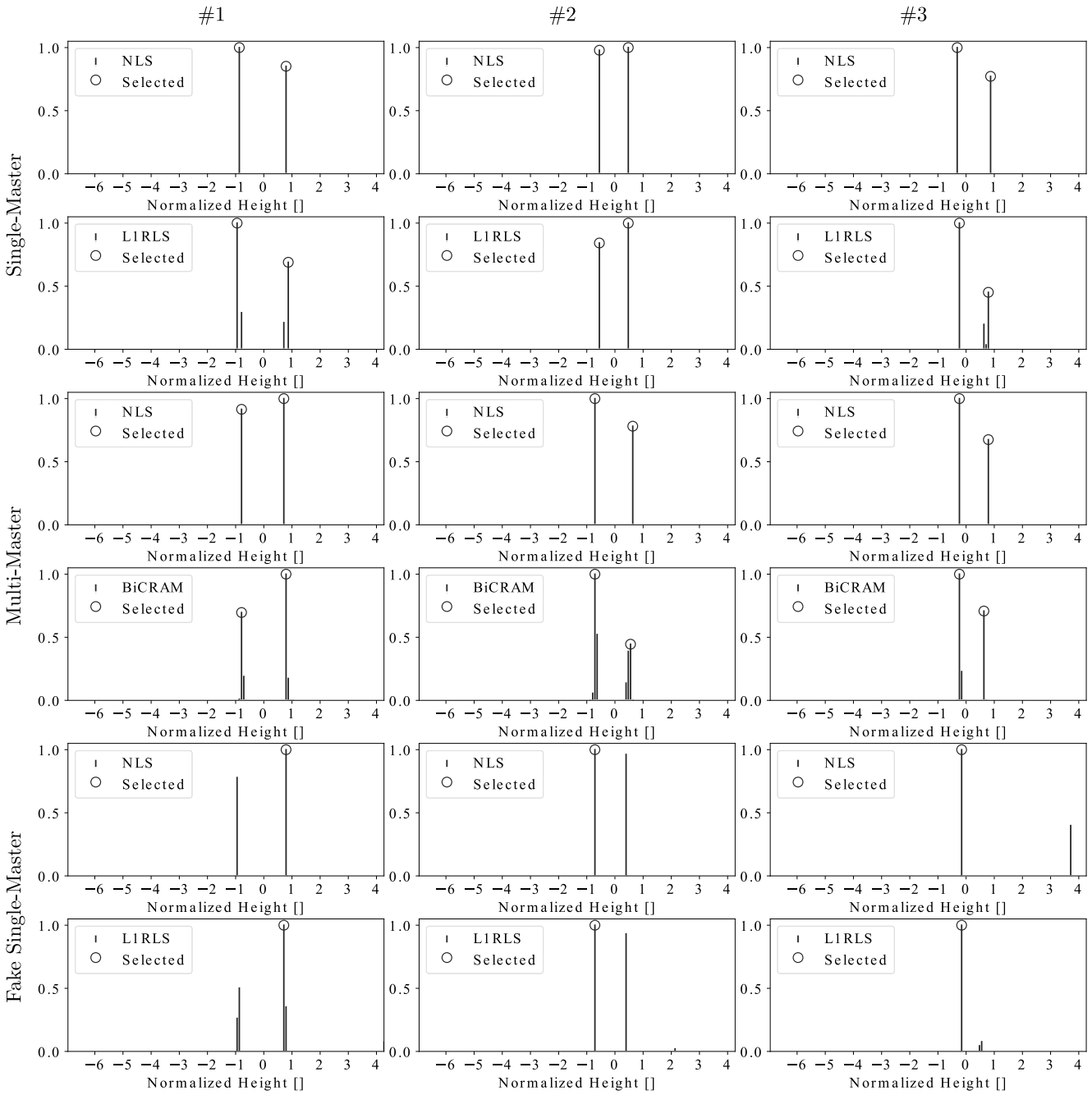


Fig. 10. Height profile estimate of #1–3 in Fig. 9. Vertical line: before model-order selection. Circle: after model-order selection.

RaySAR simulator [71] was employed to simulate the dominant scatterers that, as already mentioned in Section V-A, correspond to the triple reflections on the building facade. With the help of atmospheric and geodynamic corrections from the DLR's SAR Geodetic Processor [72], [73] and the newly enhanced TerraSAR-X orbit products [74], their absolute coordinates were converted into azimuth timing, range timing, and height that we refer to as the *Level-0* ground-truth data.

Level-1 ground-truth data consist of height at 30 simulated PSs that are matched with the real ones. The matching was conducted in the azimuth-range geometry, so as not to be affected by any height estimate error [75]. Fig. 8 (top) shows the height simulations at the subpixel azimuth-range positions of the corresponding 30 PSs. This height is relative to a corner reflector that is located on the top of a neighboring TUM building and next to a permanent GPS station [68].

In addition, we performed height interpolation for a total of 594 looks (see Section V-A) in the following way. First, the height of each simulated PS was converted into interferometric phase. Next, the distance to the polyline representing the nearest range cross section of the building facade was used as the independent variable to construct a 1-D interpolator. In the end, the phase was interpolated at the previously mentioned 594 looks and converted back into height. This interpolated height is referred to as the *Level*-2 ground truth and shown in Fig. 8 (bottom). Needless to

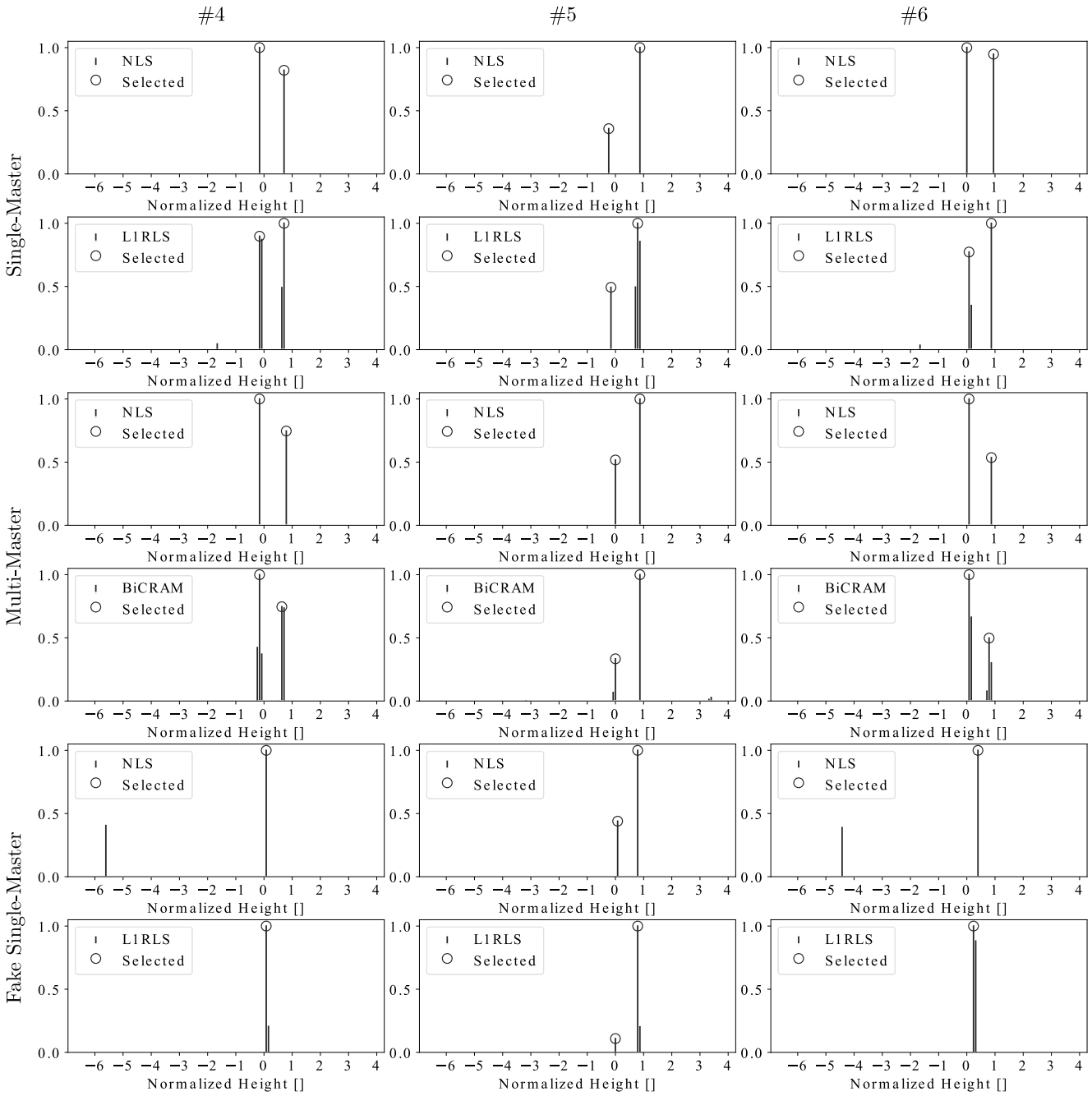


Fig. 11. Height profile estimate of #4–6 in Fig. 9. Vertical line: before model-order selection. Circle: after model-order selection.

say, one assumption is that each scatterer, if it does exist, should lie on the building facade. A cross-validation of this 1-D interpolator was performed in [68], where the standard deviation (SD) and the median absolute deviation (MAD) were shown to be 0.004 and 0.002 m, respectively.

In the next section, our preliminary results are reported.

C. Experimental Results

The experiments can be divided into three categories: single-master, multi-master, and fake single-master (see Section V-A). In each category, two algorithms were applied for tomographic reconstruction.

As a proof of concept, we selected six looks that are very likely subject to the facade-roof layover. These six looks were chosen in a systematic way: we performed tomographic reconstruction on the *single-master* stack by using Tikhonov regularization (i.e., the ℓ_1 norm in the regularization term of (36) is replaced by the ℓ_2 norm), extracted all the seven looks containing double scatterers, and discarded one look whose height distance is almost identical to the one of another look. These six looks are shown in Fig. 9, where the indices increase with the decreasing estimated height distance from approximately 1.5 to 0.8 times the vertical Rayleigh resolution. This ordering agrees approximately with intuition under the assumption that the roof is entirely flat: the higher the scatterer on the facade, the less the height distance to the roof.

The estimated height profile is shown in Figs. 10 (#1–3) and 11 (#4–6), where we used the vertical Rayleigh resolution

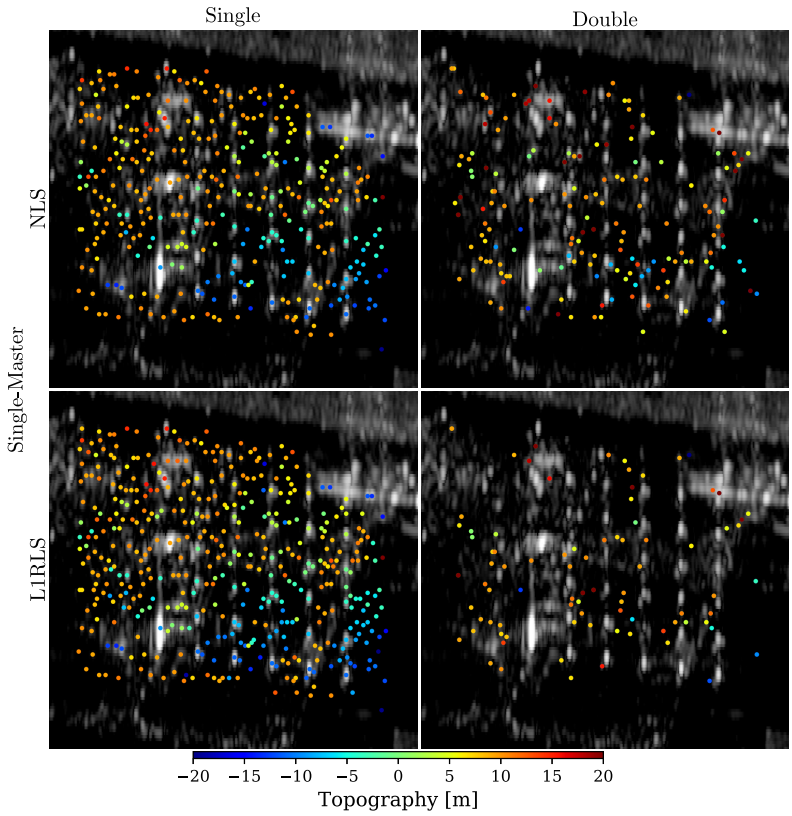


Fig. 12. Single-master height estimates of single and double scatterers. (Top) NLS. (Bottom) L1RLS.

TABLE III
SINGLE- AND MULTI-MASTER HEIGHT ESTIMATES OF SIX LAYOVER CASES [M]

		#1		#2		#3		#4		#5		#6	
		h_1	h_2	h_1	h_2	h_1	h_2	h_1	h_2	h_1	h_2	h_1	h_2
RaySAR		−9.22	–	−8.91	–	−2.39	–	−2.87	–	−1.59	–	0.55	–
Single-Master	NLS	−10.27	9.10	−8.14	6.97	−4.27	10.72	−2.63	9.97	−1.95	10.03	0.98	11.34
	L1RLS	−10.27	9.10	−8.14	6.97	−4.27	10.72	−2.63	9.97	−1.95	10.03	0.98	11.34
Multi-Master	NLS	−10.15	9.47	−8.96	8.23	−3.50	9.98	−2.06	10.45	−0.37	10.90	0.66	10.85
	BiCRAM	−10.15	9.47	−8.96	8.23	−3.50	9.98	−2.06	10.45	−0.37	10.90	0.66	10.85
Fake Single-Master	NLS	9.35	–	−9.12	–	−1.64	–	1.19	–	0.63	10.03	3.47	–
	L1RLS	9.35	–	−9.12	–	−1.64	–	1.19	–	0.63	10.03	3.47	–

TABLE IV
SINGLE- AND MULTI-MASTER RUNTIME

		Runtime [s]
Single-Master	NLS	6154
	L1RLS	736
Multi-Master	NLS	460
	BiCRAM	6853
Fake Single-Master	NLS	48
	L1RLS	65

TABLE V
SINGLE- AND MULTI-MASTER NUMBER OF SCATTERERS

		Single	Double	Ratio	Facade
Single-Master	NLS	359	332	1.08	148
	L1RLS	446	168	2.65	189
Multi-Master	NLS	260	158	1.65	124
	BiCRAM	291	118	2.47	133
Fake Single-Master	NLS	360	60	6.00	134
	L1RLS	381	38	10.03	143

of the *single-master* stack (see Section V-A) for normalizing the x-axis. The height estimates are listed in Table III. In the *single-master* setting, NLS and L1RLS produced very similar height profiles, despite the occasional sporadic artifacts in the latter, which are known to be an intrinsic problem of ℓ_1-regularization. Moreover, the height estimates were identical after off-grid correction. In each case, the height estimate of the lower scatterer fits very well the Level-2 RaySAR simulation of facade. Overall, the *multi-master* results are consistent with the *single-master* ones, with deviations of height estimates typically of several decimeters. In the *fake single-master* setting, however, layover separation was only

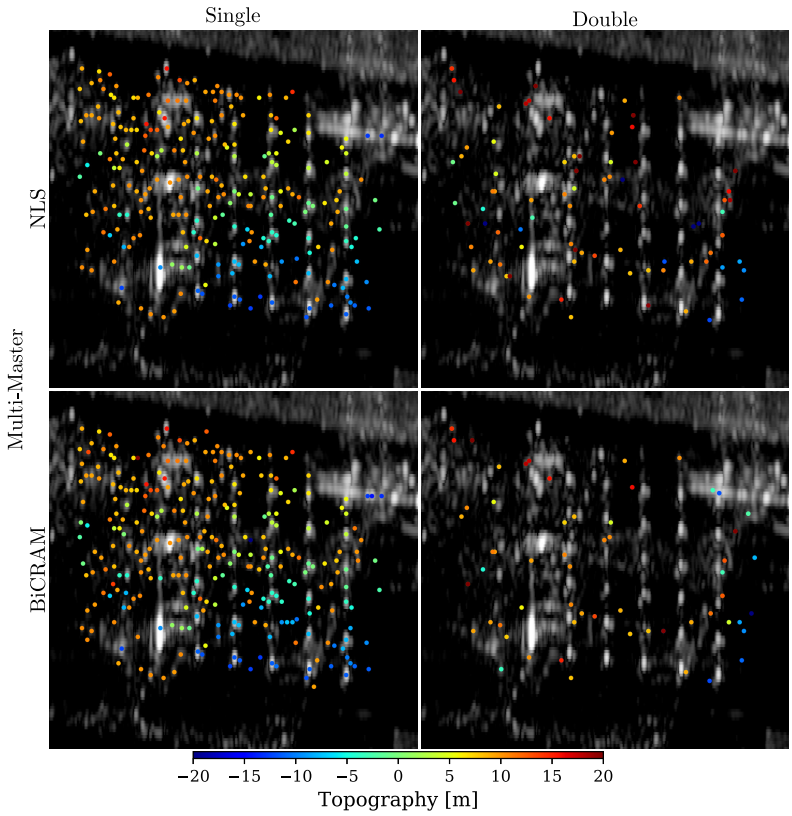


Fig. 13. Multi-master height estimates of the single and double scatterers. (Top) NLS. (Bottom) BiCRAM.

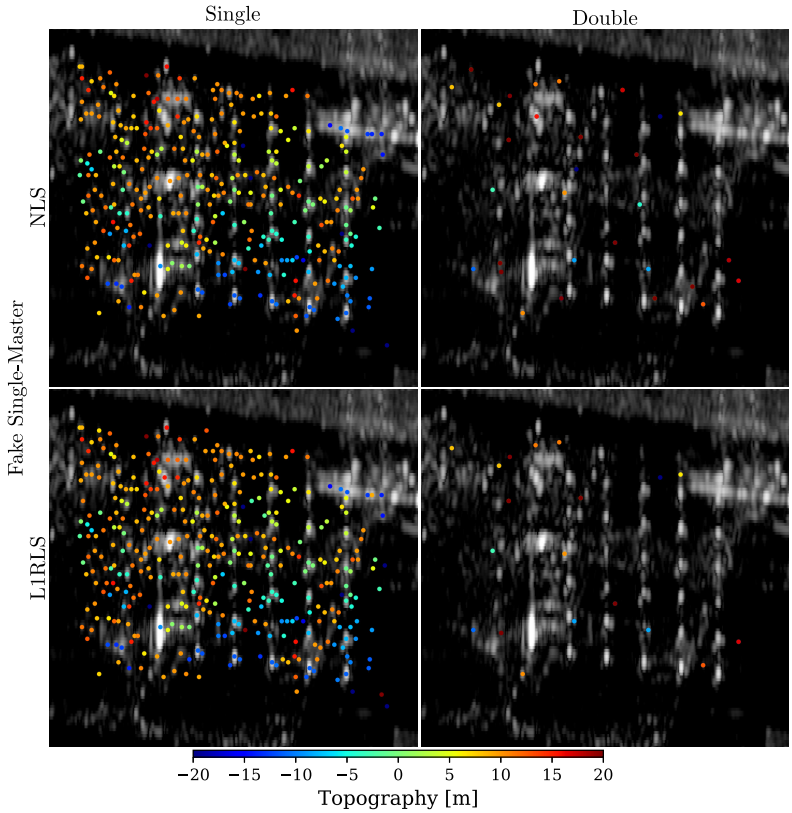


Fig. 14. Fake single-master height estimates of the single and double scatterers. (Top) NLS. (Bottom) L1RLS.

successful in the fifth case, presumably due to the high SNR (see the brightness of the look in Fig. 9). When the height distance is significantly larger than the vertical Rayleigh resolution (#1 and 2), both NLS and L1RLS could reconstruct

TABLE VI
STATISTICS OF HEIGHT ESTIMATE ERROR [M]: 30 PSS (LEVEL 1)

		Min	Max	Mean	Median	SD	MAD
Single-Master	NLS	−0.81	0.65	−0.30	−0.29	0.34	0.35
	L1RLS	−0.81	0.65	−0.31	−0.29	0.34	0.37
Multi-Master	NLS	−0.88	0.55	−0.33	−0.36	0.31	0.28
	BiCRAM	−0.88	0.55	−0.33	−0.36	0.31	0.28
Fake Single-Master	NLS	−1.02	0.68	−0.43	−0.48	0.34	0.30
	L1RLS	−1.02	0.68	−0.43	−0.48	0.34	0.30

TABLE VII
STATISTICS OF THE HEIGHT ESTIMATE ERROR [M]: EXTRACTED FACADE SCATTERERS (LEVEL 2)

		Min	Max	Mean	Median	SD	MAD
Single-Master	NLS	−3.03	2.99	−0.40	−0.51	1.23	0.96
	L1RLS	−3.03	2.99	−0.42	−0.49	1.25	0.99
Multi-Master	NLS	−3.04	2.71	−0.49	−0.40	1.13	0.85
	BiCRAM	−3.04	2.71	−0.50	−0.40	1.12	0.86
Fake Single-Master	NLS	−2.93	2.70	−0.38	−0.50	1.15	0.91
	L1RLS	−2.93	2.70	−0.35	−0.48	1.21	1.05

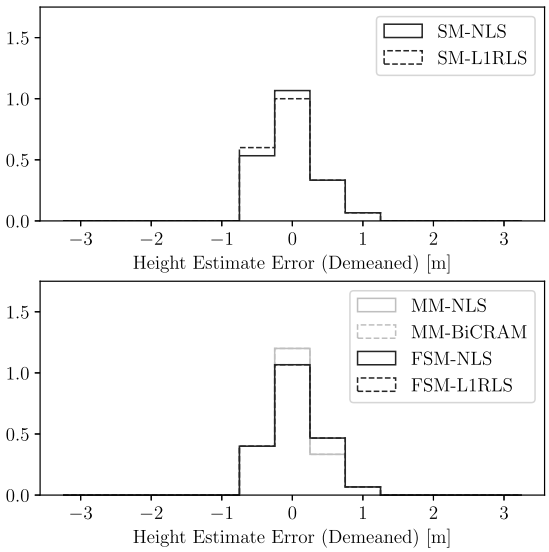


Fig. 15. Normalized histogram of the height estimate error of 30 PSs (Level 1). SM: single-master. MM: multi-master. FSM: fake single-master.

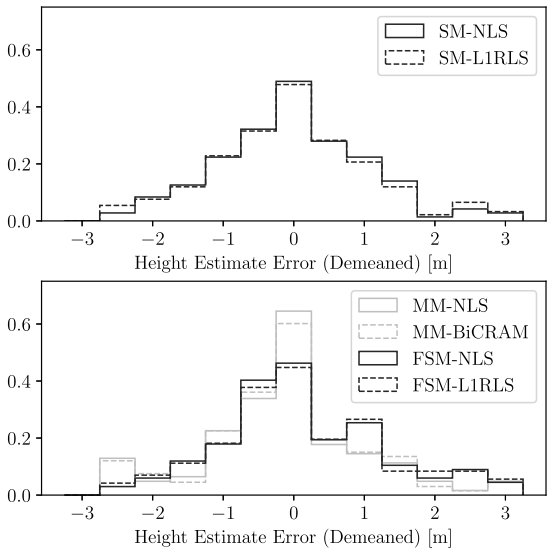


Fig. 16. Normalized histogram of the height estimate error of the extracted facade scatterers (Level 2). SM: single-master. MM: multi-master. FSM: fake single-master.

double scatterers, but only the one with larger amplitude could pass model-order selection. When the height distance approaches the vertical Rayleigh resolution or becomes even smaller (#3, 4, and 6), neither algorithm could reconstruct a second scatterer, and the height estimate of the single scatterer after off-grid correction is also arguably wrong. We are, therefore, convinced by this simple experiment that the conventional *single-master* approach, if applied to a *multi-master* stack, can be insufficient for layover separation.

Naturally, we also performed tomographic reconstruction for all the 594 looks within the building footprint in Fig. 8 (bottom). Table IV lists the overall runtime on a desktop with a quad-core Intel processor at 3.40 GHz and 16-GB RAM. Note that the periodical deformation model was only used in the *single-master* case, and the solution path of L1RLS or BiCRAM was sampled 11 times (see Section V-A). The height estimates of the single and double scatterers are shown in Figs. 12–14 for the three categories, respectively. In the case of double scatterers, the higher one was plotted. The seemingly messy appearance in the left column is due to the fact that single scatterers originate from both facade and roof. In spite of this, the gradual color transition at the 30 PSs from far range to near range agrees visually very well with the Level-1 ground truth in Fig. 8 (top). Table V lists the number of scatterers in each case. In the *single-master* setting, NLS detected almost twice as many double scatterers as L1RLS. This is presumably due to a higher false-positive rate: 2 out of 30 PS (fifth/second row from the near range and fifth/fifth column from the

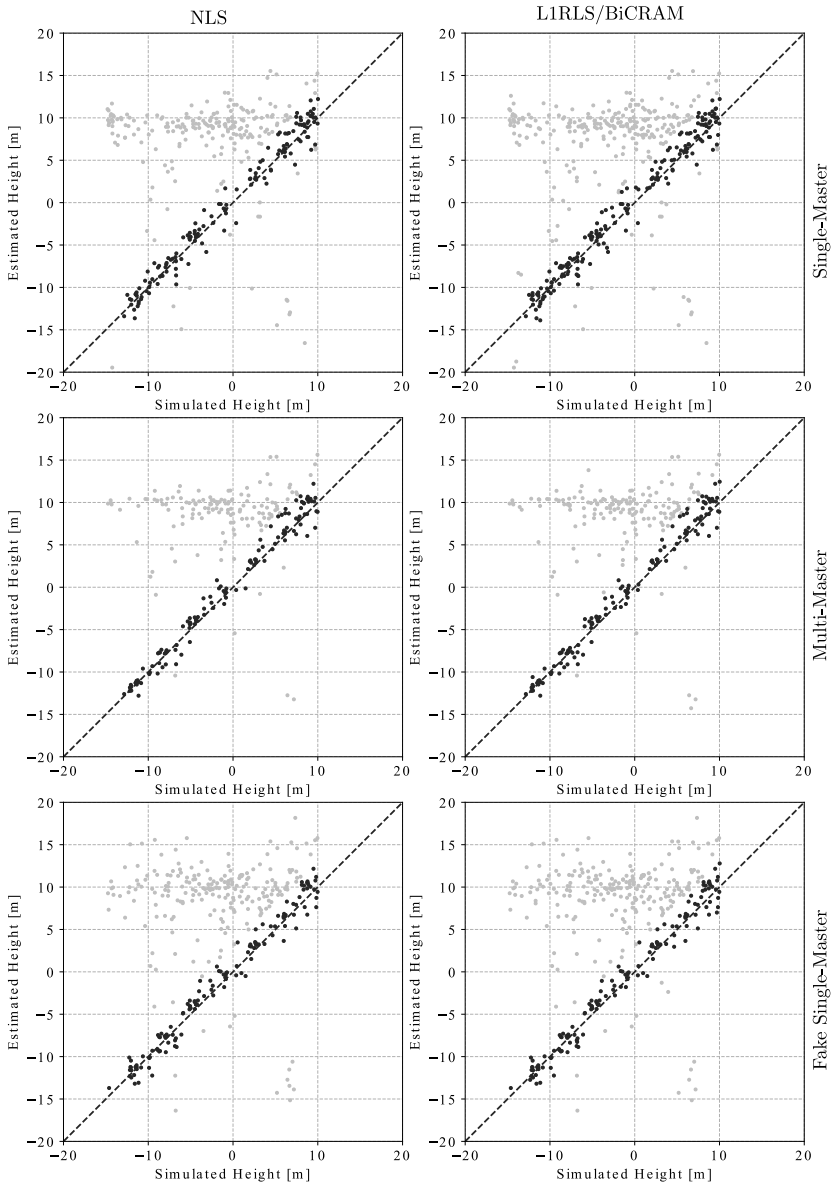


Fig. 17. Scatter plot of the simulated and estimated heights of the single scatterers using Level-2 height ground truth. Black: extracted facade scatterers. Gray: extracted nonfacade scatterers.

late azimuth on the 6×5 regular grid) double scatterers were detected, although there should only be single ones. The number of double scatterers in the *multi-master* case is in the same order as the *single-master* L1RLS, and the ratio between the number of single and the one of double scatterers is also similar. We attribute the smaller number of single scatterers to the nonconvexity of the optimization problem. In particular, as Theorem 2(2) suggests, a certain condition needs to be fulfilled for any nonzero solution of the height profile estimate to exist at all, let alone whether an algorithm can provably recover it. In the *fake single-master* category, many fewer double scatterers were produced. This is presumably due to the double scatterers being misdetected as single scatterers, which occurred five out of six times in the previous experiment (see Figs. 10 and 11).

The next section elucidates how we validated the height estimates with the Level-1 and Level-2 ground-truth data.

D. Validation

Since the height ground truth is limited to facade only (see Section V-B), the validation was focused on single scatterers by the following two approaches: the first one uses 30 PSs and the second one is based on the extracted facade scatterers.

As already mentioned in Section V-A, the 30 PSs constituting the Level-1 ground truth in Fig. 8 (top) are caused by triple reflections on the building facade and are located on a regular grid of salient points. Due to the (almost) identical scattering geometry, these PSs should have similar SNRs and are, therefore, ideal for height estimate validation. In each of the six cases, single scatterers were correctly detected at all the 30 PSs—with the exception that double scatterers were misdetected by the NLS in the *single-master* setting (see Section V-C). For this reason, the height estimate error could be evaluated straightforwardly. Fig. 15 shows the normalized histogram and Table VI lists some of its statistical parameters.

As a reference, the SD and MAD of the height estimate error of the PSI result are about 0.28 and 0.22 m, respectively [68]. In each of the three settings (single-master, multi-master, and fake single-master), the respective two algorithms performed similarly and no significant difference is visible. A cross-comparison between the multi-master and fake single-master cases revealed the superiority of the former: its histogram is more centered around zero, and both of its SD and MAD are slightly smaller. This is unsurprising, since we already analyzed the implications of the single-look *multi-master* data model for the single scatterers in Section II-B. At this point, we could confidently assert that it does make a difference in practice, albeit small, despite the longer (by approximately one order considering that the solution path of BiCRAM was sampled 11 times) processing time. Somewhat surprisingly, the *multi-master* result is also slightly better than the *single-master* one. We suspect that this is due to the complication of the *single-master* tomographic processing by using the (imperfect) periodical deformation model and the justified simplification in the *multi-master* case thanks to the small-baseline configuration (so that the deformation-induced phase is mitigated by forming interferograms).

The second approach is based on all facade scatterers (Level-2 ground truth) in Fig. 8 (bottom), given that they do exist. The scatter plots of the simulated and estimated heights of the single scatterers are shown in Fig. 17. It is obvious that many single scatterers are located on the building roof (see the gray dots above the diagonal line). In order to extract the facade scatterers, we used a threshold of $\pm 3\times$CRLB added to the simulated value. The extracted facade scatterers, whose number is given in Table V for each case, are shown as black dots and were used for height estimate validation. The normalized histogram is shown in Fig. 16, and some of its statistical parameters are provided in Table VII. Likewise, the two respective algorithms performed similarly in each setting, and the *multi-master* height estimate error has slightly less deviation. The SD and MAD are worse than those in Table VII due to the much larger range of SNRs.

The next section concludes this article and suggests some prospective work.

VI. Conclusion and Discussion

The previous sections provided new insights into the single-look multi-master SAR tomography. The single-look multi-master data model was established, and two algorithms were developed within a common inversion framework. The first algorithm extends the conventional NLS to the single-look multi-master data model, and the second one uses bi-convex relaxation and alternating minimization. Extensive efforts were devoted to studying the nonconvex objective function of the NLS subproblem and to experimenting with different acceleration techniques for the ADMM-based algorithms. We demonstrated with the help of a real TerraSAR-X data set that the conventional *single-master* approach, if applied to a *multi-master* stack, can be insufficient for layover separation, even when the height distance between the two scatterers is significantly larger than the vertical Rayleigh resolution. By means of an SAR imaging geodesy and simulation framework, we managed to generate two levels of height ground truth. The height estimates in each of the three settings were validated at either 30 PSs or hundreds of extracted facade scatterers. Overall, the *multi-master* approach performed slightly better, although it was computationally more demanding.

A special case of the general problem analyzed so far is single-look SAR tomography using only bistatic (or pursuit monostatic) interferograms. On the one hand, the advantages are that bistatic interferograms are (almost) APS-free, and the data model is still linear for any single scatterer whose reflectivity can be estimated up to a constant phase angle (7). On the other hand, the disadvantages are that, for double or multiple scatterers, bistatic interferograms are not motion-free (see Section II-B), and the data model is nonlinear (9).

An alternative way to formulate the problem in the bistatic setting is to parameterize APS without forming any interferogram. Let $\mathbf{g}$ and $\mathbf{h}$ be the bistatic observations of the master and slave scenes, respectively. We have essentially two data sets

$$\mathbf{g} \approx (\mathbf{R}\boldsymbol{\gamma}) \circ \exp(j\boldsymbol{\phi}), \quad \mathbf{h} \approx (\mathbf{S}\boldsymbol{\gamma}) \circ \exp(j\boldsymbol{\phi}) \tag{50}$$

where $\boldsymbol{\phi} \in \mathbb{R}^{N'}$ is the APS vector. Under the sparsity or compressibility assumption of $\boldsymbol{\gamma}$, one could consider the following problem:

$$\min_{\boldsymbol{\gamma},\boldsymbol{\phi}} \ \frac{1}{2}\|(\mathbf{R}\boldsymbol{\gamma}) \circ \exp(j\boldsymbol{\phi}) - \mathbf{g}\|_2^2 + \frac{1}{2}\|(\mathbf{S}\boldsymbol{\gamma}) \circ \exp(j\boldsymbol{\phi}) - \mathbf{h}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1 \tag{51}$$

or more compactly

$$\min_{\boldsymbol{\gamma},\boldsymbol{\phi}} \frac{1}{2}\|\left(\tilde{\mathbf{R}}\boldsymbol{\gamma}\right) \circ \left(\tilde{\mathbf{I}}\exp(j\boldsymbol{\phi})\right) - \tilde{\mathbf{g}}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1 \tag{52}$$

where $\tilde{\mathbf{R}} \coloneqq \begin{pmatrix}\mathbf{R}\\ \mathbf{S}\end{pmatrix}$, $\tilde{\mathbf{I}} \coloneqq \begin{pmatrix}\mathbf{I}\\ \mathbf{I}\end{pmatrix}$, and $\tilde{\mathbf{g}} \coloneqq \begin{pmatrix}\mathbf{g}\\ \mathbf{h}\end{pmatrix}$. Note that this problem is bi-convex in $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$.

Inspired by PSI, another related problem is SAR tomography on the edges. Let $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ represent the reflectivity profiles of the two neighboring looks, and their phase-calibrated SLC measurements be denoted as $\mathbf{g}$ and $\mathbf{h}$, respectively. Consider the following problem:

$$\min_{\boldsymbol{\gamma},\boldsymbol{\theta}} \frac{1}{2}\|(\mathbf{R}\boldsymbol{\gamma}) \circ \left(\overline{\mathbf{S}\boldsymbol{\theta}}\right) - \mathbf{g} \circ \overline{\mathbf{h}}\|_2^2 + \lambda_1\|\boldsymbol{\gamma}\|_1 + \lambda_2\|\boldsymbol{\theta}\|_1 \tag{53}$$

which is bi-convex in $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$. Likewise, the rationale of $\mathbf{g} \circ \overline{\mathbf{h}}$ is to mitigate APS for the neighboring looks. Alternatively, a parametric approach similar to (51) could also be considered.

Appendix A
Recap of ADMM

The ADMM [53] solves a minimization problem in the form of

$$\min_{\mathbf{x},\mathbf{z}} \ f(\mathbf{x}) + g(\mathbf{z}) \quad \text{s. t. } \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{z} = \mathbf{e} \tag{54}$$

by alternatively minimizing its augmented Lagrangian [54, p. 509]

$$L_\rho(\mathbf{x},\mathbf{y},\mathbf{z}) \coloneqq f(\mathbf{x}) + g(\mathbf{z}) + \Re\langle \mathbf{y}, \mathbf{C}\mathbf{x}+\mathbf{D}\mathbf{z}-\mathbf{e}\rangle + (\rho/2)\|\mathbf{C}\mathbf{x}+\mathbf{D}\mathbf{z}-\mathbf{e}\|_2^2 \quad (55)$$

that is

$$\begin{aligned}\mathbf{x}^{(k+1)} &\coloneqq \arg\min_{\mathbf{x}} L_\rho\big(\mathbf{x},\mathbf{y}^{(k)},\mathbf{z}^{(k)}\big)\\ \mathbf{z}^{(k+1)} &\coloneqq \arg\min_{\mathbf{z}} L_\rho\big(\mathbf{x}^{(k+1)},\mathbf{y}^{(k)},\mathbf{z}\big)\\ \mathbf{y}^{(k+1)} &\coloneqq \mathbf{y}^{(k)} + \rho\big(\mathbf{C}\mathbf{x}^{(k+1)}+\mathbf{D}\mathbf{z}^{(k+1)}-\mathbf{e}\big)\end{aligned} \quad (56)$$

in the kth iteration, where $\rho \in \mathbb{R}_{++}$ is a penalty parameter.

Appendix B
Proof of Proposition 1

The proof uses the following minor result.

Lemma 6: For any $\mathbf{F},\mathbf{G} \in \mathbb{R}^{n\times n}$ and $c,d \in \mathbb{R}$ such that $c^2+d^2=1$, the following equalities hold:

$$\begin{pmatrix} c\mathbf{I} & d\mathbf{I}\\ -d\mathbf{I} & c\mathbf{I}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{F} & -\mathbf{G}\\ \mathbf{G} & \mathbf{F}\end{pmatrix}\begin{pmatrix} c\mathbf{I} & d\mathbf{I}\\ -d\mathbf{I} & c\mathbf{I}\end{pmatrix}=\begin{pmatrix}\mathbf{F} & -\mathbf{G}\\ \mathbf{G} & \mathbf{F}\end{pmatrix}$$

$$\begin{pmatrix} c\mathbf{I} & d\mathbf{I}\\ -d\mathbf{I} & c\mathbf{I}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{F} & \mathbf{G}\\ \mathbf{G} & -\mathbf{F}\end{pmatrix}=\begin{pmatrix}\mathbf{F} & \mathbf{G}\\ \mathbf{G} & -\mathbf{F}\end{pmatrix}\begin{pmatrix} c\mathbf{I} & d\mathbf{I}\\ -d\mathbf{I} & c\mathbf{I}\end{pmatrix}. \quad (57)$$

Proof: Observe that for any $a,b \in \mathbb{R}$ such that $a^2+b^2 \neq 0$

$$\begin{pmatrix} a\mathbf{I} & b\mathbf{I}\\ -b\mathbf{I} & a\mathbf{I}\end{pmatrix}^{-1}=\frac{1}{a^2+b^2}\begin{pmatrix} a\mathbf{I} & -b\mathbf{I}\\ b\mathbf{I} & a\mathbf{I}\end{pmatrix}. \quad (58)$$

The rest of the proof follows by straightforward computations. □

Now, we turn our attention to the proposition.

Proof of Proposition 1: First, we prove that $\nabla^2 f(\mathbf{x})$ and $\nabla^2 f(\mathbf{x}\exp(j\phi))$ are similar, i.e., there exists an invertible $\mathbf{P}$ such that $\nabla^2 f(\mathbf{x}) = \mathbf{P}^{-1}\nabla^2 f(\mathbf{x}\exp(j\phi))\mathbf{P}$.

Observe that

$$\begin{aligned}&\mathbf{C}(\mathbf{x}\exp(j\phi))\\ &\quad= \mathbf{A}^H \operatorname{Diag}\big((\mathbf{B}\mathbf{x}\exp(j\phi))\circ\big(\overline{\mathbf{B}\mathbf{x}\exp(j\phi)}\big)\big)\mathbf{A}\\ &\qquad+\mathbf{B}^H \operatorname{Diag}\big((\mathbf{A}\mathbf{x}\exp(j\phi))\circ\big(\overline{\mathbf{A}\mathbf{x}\exp(j\phi)}\big)\big)\mathbf{B}\\ &\quad= \mathbf{C}(\mathbf{x})\\ &\mathbf{D}(\mathbf{x}\exp(j\phi))\\ &\quad= \mathbf{A}^H \operatorname{Diag}((\mathbf{A}\mathbf{x}\exp(j\phi))\circ(\mathbf{B}\mathbf{x}\exp(j\phi)))\overline{\mathbf{B}}\\ &\qquad+\mathbf{B}^H \operatorname{Diag}((\mathbf{A}\mathbf{x}\exp(j\phi))\circ(\mathbf{B}\mathbf{x}\exp(j\phi)))\overline{\mathbf{A}}\\ &\quad= \mathbf{D}(\mathbf{x})\exp(j2\phi)\\ &\mathbf{E}(\mathbf{x}\exp(j\phi))\\ &\quad= \mathbf{A}^H \operatorname{Diag}\big((\mathbf{A}\mathbf{x}\exp(j\phi))\circ\big(\overline{\mathbf{B}\mathbf{x}\exp(j\phi)}\big)-\mathbf{b}\big)\mathbf{B}\\ &\qquad+\mathbf{B}^H \operatorname{Diag}\big(\big(\overline{\mathbf{A}\mathbf{x}\exp(j\phi)}\big)\circ(\mathbf{B}\mathbf{x}\exp(j\phi))-\overline{\mathbf{b}}\big)\mathbf{A}\\ &\quad= \mathbf{E}(\mathbf{x}).\end{aligned} \quad (59)$$

Let $\mathbf{C} \coloneqq \mathbf{C}(\mathbf{x})$, $\mathbf{D} \coloneqq \mathbf{D}(\mathbf{x})$, and $\mathbf{E} \coloneqq \mathbf{E}(\mathbf{x})$. The Hessian becomes

$$\begin{aligned}&\nabla^2 f(\mathbf{x}\exp(j\phi))\\ &\quad=\begin{pmatrix}\Re(\mathbf{C}) & -\Im(\mathbf{C})\\ \Im(\mathbf{C}) & \Re(\mathbf{C})\end{pmatrix}\\ &\qquad+\begin{pmatrix}\Re(\mathbf{D}\exp(j2\phi)) & \Im(\mathbf{D}\exp(j2\phi))\\ \Im(\mathbf{D}\exp(j2\phi)) & -\Re(\mathbf{D}\exp(j2\phi))\end{pmatrix}\\ &\qquad+\begin{pmatrix}\Re(\mathbf{E}) & -\Im(\mathbf{E})\\ \Im(\mathbf{E}) & \Re(\mathbf{E})\end{pmatrix}\\ &\quad=\begin{pmatrix}\Re(\mathbf{C}+\mathbf{E}) & -\Im(\mathbf{C}+\mathbf{E})\\ \Im(\mathbf{C}+\mathbf{E}) & \Re(\mathbf{C}+\mathbf{E})\end{pmatrix}\\ &\qquad+\begin{pmatrix}\Re(\mathbf{D}) & \Im(\mathbf{D})\\ \Im(\mathbf{D}) & -\Re(\mathbf{D})\end{pmatrix}\begin{pmatrix}\cos(2\phi)\mathbf{I} & \sin(2\phi)\mathbf{I}\\ -\sin(2\phi)\mathbf{I} & \cos(2\phi)\mathbf{I}\end{pmatrix}.\end{aligned} \quad (60)$$

The choice of $\mathbf{P}$ can be divided into two cases depending on the value of ϕ.

In the trivial case, $\phi = (2k+1)\pi/2$ for some $k \in \mathbb{Z}$. Let

$$\mathbf{P} \coloneqq \begin{pmatrix}\mathbf{0} & -\mathbf{I}\\ \mathbf{I} & \mathbf{0}\end{pmatrix}. \quad (61)$$

This leads to

$$\begin{aligned}&\mathbf{P}^{-1}\nabla^2 f(\mathbf{x}\exp(j\phi))\mathbf{P}\\ &\quad=\begin{pmatrix}\mathbf{0} & -\mathbf{I}\\ \mathbf{I} & \mathbf{0}\end{pmatrix}^{-1}\begin{pmatrix}\Re(\mathbf{C}+\mathbf{E}) & -\Im(\mathbf{C}+\mathbf{E})\\ \Im(\mathbf{C}+\mathbf{E}) & \Re(\mathbf{C}+\mathbf{E})\end{pmatrix}\begin{pmatrix}\mathbf{0} & -\mathbf{I}\\ \mathbf{I} & \mathbf{0}\end{pmatrix}\\ &\qquad+\begin{pmatrix}\mathbf{0} & -\mathbf{I}\\ \mathbf{I} & \mathbf{0}\end{pmatrix}^{-1}\begin{pmatrix}\Re(\mathbf{D}) & \Im(\mathbf{D})\\ \Im(\mathbf{D}) & -\Re(\mathbf{D})\end{pmatrix}\begin{pmatrix}-\mathbf{I} & \mathbf{0}\\ \mathbf{0} & -\mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{0} & -\mathbf{I}\\ \mathbf{I} & \mathbf{0}\end{pmatrix}\\ &\quad=\begin{pmatrix}\Re(\mathbf{C}+\mathbf{E}) & -\Im(\mathbf{C}+\mathbf{E})\\ \Im(\mathbf{C}+\mathbf{E}) & \Re(\mathbf{C}+\mathbf{E})\end{pmatrix}\\ &\qquad+\begin{pmatrix}\Re(\mathbf{D}) & \Im(\mathbf{D})\\ \Im(\mathbf{D}) & -\Re(\mathbf{D})\end{pmatrix}\begin{pmatrix}\mathbf{0} & -\mathbf{I}\\ \mathbf{I} & \mathbf{0}\end{pmatrix}\begin{pmatrix}-\mathbf{I} & \mathbf{0}\\ \mathbf{0} & -\mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{0} & -\mathbf{I}\\ \mathbf{I} & \mathbf{0}\end{pmatrix}\\ &\quad=\begin{pmatrix}\Re(\mathbf{C}+\mathbf{E}) & -\Im(\mathbf{C}+\mathbf{E})\\ \Im(\mathbf{C}+\mathbf{E}) & \Re(\mathbf{C}+\mathbf{E})\end{pmatrix}+\begin{pmatrix}\Re(\mathbf{D}) & \Im(\mathbf{D})\\ \Im(\mathbf{D}) & -\Re(\mathbf{D})\end{pmatrix}\\ &\quad=\nabla^2 f(\mathbf{x})\end{aligned} \quad (62)$$

where the second equality follows from Lemma 6.

In the nontrivial case, $\phi \neq (2k+1)\pi/2$ for any $k \in \mathbb{Z}$. Let

$$\mathbf{P} \coloneqq \begin{pmatrix}\sqrt{\dfrac{1+\cos(2\phi)}{2}}\mathbf{I} & -\dfrac{\sin(2\phi)}{\sqrt{2(1+\cos(2\phi))}}\mathbf{I}\\ \dfrac{\sin(2\phi)}{\sqrt{2(1+\cos(2\phi))}}\mathbf{I} & \sqrt{\dfrac{1+\cos(2\phi)}{2}}\mathbf{I}\end{pmatrix}. \quad (63)$$

Likewise, the same equality holds.

Finally, we use the similarity property to show that an eigenvalue of $\nabla^2 f(\mathbf{x})$ is also an eigenvalue of $\nabla^2 f(\mathbf{x}\exp(j\phi))$.

Let $(\lambda,\mathbf{v})$ be an eigenpair of $\nabla^2 f(\mathbf{x})$. The similarity property implies

$$\begin{aligned}\lambda\mathbf{v} &= \nabla^2 f(\mathbf{x})\mathbf{v} = \mathbf{P}^{-1}\nabla^2 f(\mathbf{x}\exp(j\phi))\mathbf{P}\mathbf{v}\\ &\implies \nabla^2 f(\mathbf{x}\exp(j\phi))\mathbf{P}\mathbf{v} = \lambda\mathbf{P}\mathbf{v}\end{aligned} \quad (64)$$

that is, $(\lambda,\mathbf{P}\mathbf{v})$ is an eigenpair of $\nabla^2 f(\mathbf{x}\exp(j\phi))$. The proof in the other direction is straightforward. □

APPENDIX C
PROOF OF THEOREM 2

Before we delve into the proof, it is useful to define a few auxiliary variables. Let

$$\tilde{\mathbf{C}} \coloneqq \begin{pmatrix} \Re(\mathbf{C}) & -\Im(\mathbf{C}) \\ \Im(\mathbf{C}) & \Re(\mathbf{C}) \end{pmatrix}$$
$$\tilde{\mathbf{D}} \coloneqq \begin{pmatrix} \Re(\mathbf{D}) & \Im(\mathbf{D}) \\ \Im(\mathbf{D}) & -\Re(\mathbf{D}) \end{pmatrix}$$
$$\tilde{\mathbf{E}} \coloneqq \begin{pmatrix} \Re(\mathbf{E}) & -\Im(\mathbf{E}) \\ \Im(\mathbf{E}) & \Re(\mathbf{E}) \end{pmatrix} \tag{65}$$

so that $\nabla^2 f(\mathbf{x}) = \tilde{\mathbf{C}}+\tilde{\mathbf{D}}+\tilde{\mathbf{E}}$. Likewise, $\tilde{\mathbf{C}}, \tilde{\mathbf{D}}, \tilde{\mathbf{E}} : \mathbb{C}^n \to \mathbb{R}^{2n\times 2n}$ are the de facto (composite) functions of $\mathbf{x}$. The proof of the main theorem is based on the following minor result.

Lemma 7: For any $\mathbf{x} \coloneqq \mathbf{x}_R + j\mathbf{x}_I$, the following equalities hold:

$$\begin{aligned}
\begin{pmatrix} \mathbf{x}_R^T & \mathbf{x}_I^T \end{pmatrix} \tilde{\mathbf{C}} \begin{pmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{pmatrix} &= \mathbf{x}^H \mathbf{C} \mathbf{x} \\
\begin{pmatrix} \mathbf{x}_R^T & \mathbf{x}_I^T \end{pmatrix} \tilde{\mathbf{D}} \begin{pmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{pmatrix} &= \Re\big(\mathbf{x}^H \mathbf{D} \overline{\mathbf{x}}\big) \\
\begin{pmatrix} \mathbf{x}_R^T & \mathbf{x}_I^T \end{pmatrix} \tilde{\mathbf{E}} \begin{pmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{pmatrix} &= \mathbf{x}^H \mathbf{E} \mathbf{x} \\
\tilde{\mathbf{C}} \begin{pmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{pmatrix} &= \begin{pmatrix} \Re(\mathbf{C}\mathbf{x}) \\ \Im(\mathbf{C}\mathbf{x}) \end{pmatrix} \\
\tilde{\mathbf{D}} \begin{pmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{pmatrix} &= \begin{pmatrix} \Re(\mathbf{D}\overline{\mathbf{x}}) \\ \Im(\mathbf{D}\overline{\mathbf{x}}) \end{pmatrix} \\
\tilde{\mathbf{E}} \begin{pmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{pmatrix} &= \begin{pmatrix} \Re(\mathbf{E}\mathbf{x}) \\ \Im(\mathbf{E}\mathbf{x}) \end{pmatrix}.
\end{aligned} \tag{66}$$

Proof: The proof follows by straightforward computations. □

Proof of Theorem 2: (1) Since $\nabla f(\mathbf{0}) = \mathbf{0}$, $\mathbf{0}$ is a critical point. Observe that

$$\begin{aligned}
\mathbf{C}(\mathbf{0}) &= \mathbf{D}(\mathbf{0}) = \mathbf{0} \\
\mathbf{E}(\mathbf{0}) &= -\mathbf{A}^H \operatorname{Diag}(\mathbf{b})\mathbf{B} - \mathbf{B}^H \operatorname{Diag}\big(\overline{\mathbf{b}}\big)\mathbf{A}.
\end{aligned} \tag{67}$$

For any $\mathbf{x} \coloneqq \mathbf{x}_R + j\mathbf{x}_I \neq \mathbf{0}$

$$\begin{aligned}
&\begin{pmatrix} \mathbf{x}_R^T & \mathbf{x}_I^T \end{pmatrix} \nabla^2 f(\mathbf{0}) \begin{pmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{pmatrix} \\
&\quad = \begin{pmatrix} \mathbf{x}_R^T & \mathbf{x}_I^T \end{pmatrix} \tilde{\mathbf{E}}(\mathbf{0}) \begin{pmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{pmatrix} \\
&\quad = \mathbf{x}^H \mathbf{E}(\mathbf{0}) \mathbf{x}
\end{aligned} \tag{68}$$

where the second equality is given by Lemma 7. Since $\mathbf{E}(\mathbf{0})$ is Hermitian, we have

$$\begin{cases} \mathbf{x}^H \mathbf{E}(\mathbf{0}) \mathbf{x} > 0, & \text{if } \mathbf{A}^H \operatorname{Diag}(\mathbf{b})\mathbf{B} + \mathbf{B}^H \operatorname{Diag}\big(\overline{\mathbf{b}}\big)\mathbf{A} \prec \mathbf{0} \\ \mathbf{x}^H \mathbf{E}(\mathbf{0}) \mathbf{x} < 0, & \text{if } \mathbf{A}^H \operatorname{Diag}(\mathbf{b})\mathbf{B} + \mathbf{B}^H \operatorname{Diag}\big(\overline{\mathbf{b}}\big)\mathbf{A} \succ \mathbf{0} \end{cases} \tag{69}$$

for any $\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$.

(2) Suppose $\exists \mathbf{z} \neq \mathbf{0}$ such that $\nabla f(\mathbf{z}) = \mathbf{0}$. Equation (26) implies

$$\begin{aligned}
\mathbf{d}(\mathbf{z}) &= \mathbf{A}^H \Big(\big((\mathbf{A}\mathbf{z}) \circ \big(\overline{\mathbf{B}\mathbf{z}}\big) - \mathbf{b} \big) \circ (\mathbf{B}\mathbf{z}) \Big) \\
&\quad + \mathbf{B}^H \Big(\big(\big(\overline{\mathbf{A}\mathbf{z}}\big) \circ (\mathbf{B}\mathbf{z}) - \overline{\mathbf{b}} \big) \circ (\mathbf{A}\mathbf{z}) \Big) \\
&= \mathbf{0}.
\end{aligned} \tag{70}$$

A few manipulations lead to

$$\begin{aligned}
&\mathbf{A}^H \big((\mathbf{A}\mathbf{z}) \circ \big(\overline{\mathbf{B}\mathbf{z}}\big) \circ (\mathbf{B}\mathbf{z}) \big) + \mathbf{B}^H \big(\big(\overline{\mathbf{A}\mathbf{z}}\big) \circ (\mathbf{B}\mathbf{z}) \circ (\mathbf{A}\mathbf{z}) \big) \\
&\qquad = \mathbf{A}^H \operatorname{Diag}(\mathbf{b})\mathbf{B}\mathbf{z} + \mathbf{B}^H \operatorname{Diag}\big(\overline{\mathbf{b}}\big)\mathbf{A}\mathbf{z}.
\end{aligned} \tag{71}$$

Multiplying both sides with $\mathbf{z}^H$ on the left-hand side yields

$$\mathbf{z}^H \big(\mathbf{A}^H \operatorname{Diag}(\mathbf{b})\mathbf{B} + \mathbf{B}^H \operatorname{Diag}\big(\overline{\mathbf{b}}\big)\mathbf{A} \big)\mathbf{z} = 2\|(\mathbf{A}\mathbf{z}) \circ \big(\overline{\mathbf{B}\mathbf{z}}\big)\|_2^2 \geq 0 \tag{72}$$

which implies $\mathbf{A}^H \operatorname{Diag}(\mathbf{b})\mathbf{B} + \mathbf{B}^H \operatorname{Diag}(\overline{\mathbf{b}})\mathbf{A} \not\prec \mathbf{0}$.

(3a) We prove that $\nabla^2 f(\mathbf{z})$ is rank-deficient by showing $\nabla^2 f(\mathbf{z}) \begin{pmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{pmatrix} = \mathbf{0}$, where $\mathbf{x} \coloneqq \mathbf{z}_I - j\mathbf{z}_R$. Observe the following.

1) $(\mathbf{A}\mathbf{x}) \circ \big(\overline{\mathbf{B}\mathbf{z}}\big) + (\mathbf{A}\mathbf{z}) \circ \big(\overline{\mathbf{B}\mathbf{x}}\big) = \mathbf{0}$.
2) For any $\mathbf{F}$, $\mathbf{F}\mathbf{z} = \mathbf{0} \implies \mathbf{F}\mathbf{x} = \mathbf{0}$, which together with (70) implies

$$\begin{aligned}
\mathbf{0} = \mathbf{A}^H &\operatorname{Diag}\big((\mathbf{A}\mathbf{z}) \circ \big(\overline{\mathbf{B}\mathbf{z}}\big) - \mathbf{b} \big)\mathbf{B}\mathbf{x} \\
&+ \mathbf{B}^H \operatorname{Diag}\big(\big(\overline{\mathbf{A}\mathbf{z}}\big) \circ (\mathbf{B}\mathbf{z}) - \overline{\mathbf{b}} \big)\mathbf{A}\mathbf{x}.
\end{aligned} \tag{73}$$

By Lemma 7, it suffices to check $\mathbf{C}\mathbf{x} + \mathbf{D}\overline{\mathbf{x}} + \mathbf{E}\mathbf{x}$

$$\begin{aligned}
&\mathbf{C}\mathbf{x} + \mathbf{D}\overline{\mathbf{x}} + \mathbf{E}\mathbf{x} \\
&= \mathbf{A}^H \operatorname{Diag}\big((\mathbf{B}\mathbf{z}) \circ \big(\overline{\mathbf{B}\mathbf{z}}\big) \big)\mathbf{A}\mathbf{x} \\
&\quad + \mathbf{B}^H \operatorname{Diag}\big((\mathbf{A}\mathbf{z}) \circ \big(\overline{\mathbf{A}\mathbf{z}}\big) \big)\mathbf{B}\mathbf{x} \\
&\quad + \mathbf{A}^H \operatorname{Diag}\big((\mathbf{A}\mathbf{z}) \circ (\mathbf{B}\mathbf{z}) \big)\overline{\mathbf{B}\mathbf{x}} \\
&\quad + \mathbf{B}^H \operatorname{Diag}\big((\mathbf{A}\mathbf{z}) \circ (\mathbf{B}\mathbf{z}) \big)\overline{\mathbf{A}\mathbf{x}} \\
&\quad + \mathbf{A}^H \operatorname{Diag}\big((\mathbf{A}\mathbf{z}) \circ \big(\overline{\mathbf{B}\mathbf{z}}\big) - \mathbf{b} \big)\mathbf{B}\mathbf{x} \\
&\quad + \mathbf{B}^H \operatorname{Diag}\big(\big(\overline{\mathbf{A}\mathbf{z}}\big) \circ (\mathbf{B}\mathbf{z}) - \overline{\mathbf{b}} \big)\mathbf{A}\mathbf{x} \\
&= \mathbf{A}^H \operatorname{Diag}(\mathbf{B}\mathbf{z}) \big((\mathbf{A}\mathbf{x}) \circ \big(\overline{\mathbf{B}\mathbf{z}}\big) + (\mathbf{A}\mathbf{z}) \circ \big(\overline{\mathbf{B}\mathbf{x}}\big) \big) \\
&\quad + \mathbf{B}^H \operatorname{Diag}(\mathbf{A}\mathbf{z}) \big(\big(\overline{\mathbf{A}\mathbf{z}}\big) \circ (\mathbf{B}\mathbf{x}) + \big(\overline{\mathbf{A}\mathbf{x}}\big) \circ (\mathbf{B}\mathbf{z}) \big) \\
&= \mathbf{0}.
\end{aligned} \tag{74}$$

(3b) For any $\phi \in \mathbb{R}$, it is obvious that

$$\begin{aligned}
\mathbf{d}(\mathbf{z}\exp(j\phi)) &= \mathbf{d}(\mathbf{z}) = \mathbf{0} \\
f(\mathbf{z}\exp(j\phi)) &= f(\mathbf{z}),
\end{aligned} \tag{75}$$

that is, $\mathbf{z}\exp(j\phi)$ is also a nonzero critical point that is as good. By Proposition 1, $\nabla^2 f(\mathbf{x}\exp(j\phi))$ has the same eigenvalues as $\nabla^2 f(\mathbf{x})$ and, therefore, the same definiteness. □

APPENDIX D
PROOF OF THEOREM 3

Without loss of generality, assume that $\|\mathbf{A} \circ \overline{\mathbf{B}}\| \neq 0$.

Proof of Theorem 3: (1) When $n = 1$, observe that

$$\mathbf{A}^H \operatorname{Diag}(\mathbf{b})\mathbf{B} + \mathbf{B}^H \operatorname{Diag}\big(\overline{\mathbf{b}}\big)\mathbf{A} = 2\,\Re\Big(\big(\mathbf{A} \circ \overline{\mathbf{B}}\big)^H \mathbf{b} \Big). \tag{76}$$

The rest follows directly from Theorem 2(1).

(2) For any $\mathbf{z} \in \mathbb{C} \setminus \{0\}$, we have

$$\begin{aligned}
\mathbf{d}(\mathbf{z}) &= \mathbf{A}^H \Big(\big((\mathbf{A}\mathbf{z}) \circ \big(\overline{\mathbf{B}\mathbf{z}}\big) - \mathbf{b} \big) \circ (\mathbf{B}\mathbf{z}) \Big) \\
&\quad + \mathbf{B}^H \Big(\big(\big(\overline{\mathbf{A}\mathbf{z}}\big) \circ (\mathbf{B}\mathbf{z}) - \overline{\mathbf{b}} \big) \circ (\mathbf{A}\mathbf{z}) \Big) \\
&= \mathbf{z}|\mathbf{z}|^2 \mathbf{A}^H \big(\mathbf{A} \circ \overline{\mathbf{B}} \circ \mathbf{B} \big) - \mathbf{z}\mathbf{A}^H (\mathbf{b} \circ \mathbf{B}) \\
&\quad + \mathbf{z}|\mathbf{z}|^2 \mathbf{B}^H \big(\overline{\mathbf{A}} \circ \mathbf{B} \circ \mathbf{A} \big) - \mathbf{z}\mathbf{B}^H \big(\overline{\mathbf{b}} \circ \mathbf{A} \big)
\end{aligned}$$

$$= \mathbf{z}|\mathbf{z}|^2\|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2 - \mathbf{z}\big(\mathbf{A}\circ\overline{\mathbf{B}}\big)^H\mathbf{b} + \mathbf{z}|\mathbf{z}|^2\|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2 - \mathbf{z}\big(\overline{\mathbf{A}}\circ\mathbf{B}\big)^H\overline{\mathbf{b}} = 2\mathbf{z}\Big(\|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2|\mathbf{z}|^2 - \Re\Big(\big(\mathbf{A}\circ\overline{\mathbf{B}}\big)^H\mathbf{b}\Big)\Big) \quad (77)$$

which has a nonzero root if and only if $\Re((\mathbf{A}\circ\overline{\mathbf{B}})^H\mathbf{b}) > 0$. If this condition is satisfied, its power is given by

$$|\mathbf{z}|^2 = \Re\Big(\big(\mathbf{A}\circ\overline{\mathbf{B}}\big)^H\mathbf{b}\Big)/\|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2. \quad (78)$$

(3a) Suppose $\exists\mathbf{z}\neq\mathbf{0}$ such that $\nabla f(\mathbf{z})=\mathbf{0}$. Equations (26) and (2) imply

$$\|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2|\mathbf{z}|^2 - \Re\Big(\big(\mathbf{A}\circ\overline{\mathbf{B}}\big)^H\mathbf{b}\Big) = \mathbf{0}. \quad (79)$$

By Lemma 7, we have for any $\mathbf{x} \coloneqq \mathbf{x}_R + j\mathbf{x}_I \neq 0$

$$\begin{aligned}
&\begin{pmatrix}\mathbf{x}_R^T & \mathbf{x}_I^T\end{pmatrix}\nabla^2 f(\mathbf{z})\begin{pmatrix}\mathbf{x}_R\\ \mathbf{x}_I\end{pmatrix}\\
&= \mathbf{x}^H\mathbf{C}(\mathbf{z})\mathbf{x} + \Re\big(\mathbf{x}^H\mathbf{D}(\mathbf{z})\overline{\mathbf{x}}\big) + \mathbf{x}^H\mathbf{E}(\mathbf{z})\mathbf{x}\\
&= |\mathbf{x}|^2\mathbf{A}^H\operatorname{Diag}\big((\mathbf{B}\mathbf{z})\circ\big(\overline{\mathbf{B}\mathbf{z}}\big)\big)\mathbf{A}\\
&\quad + |\mathbf{x}|^2\mathbf{B}^H\operatorname{Diag}\big((\mathbf{A}\mathbf{z})\circ\big(\overline{\mathbf{A}\mathbf{z}}\big)\big)\mathbf{B}\\
&\quad + \Re\big(\overline{\mathbf{x}}^2\mathbf{A}^H\operatorname{Diag}((\mathbf{A}\mathbf{z})\circ(\mathbf{B}\mathbf{z}))\overline{\mathbf{B}}\big)\\
&\quad + \Re\big(\overline{\mathbf{x}}^2\mathbf{B}^H\operatorname{Diag}((\mathbf{A}\mathbf{z})\circ(\mathbf{B}\mathbf{z}))\overline{\mathbf{A}}\big)\\
&\quad + |\mathbf{x}|^2\mathbf{A}^H\operatorname{Diag}\big((\mathbf{A}\mathbf{z})\circ\big(\overline{\mathbf{B}\mathbf{z}}\big) - \mathbf{b}\big)\mathbf{B}\\
&\quad + |\mathbf{x}|^2\mathbf{B}^H\operatorname{Diag}\big(\big(\overline{\mathbf{A}\mathbf{z}}\big)\circ(\mathbf{B}\mathbf{z}) - \overline{\mathbf{b}}\big)\mathbf{A}\\
&= 2\|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2|\mathbf{x}|^2|\mathbf{z}|^2 + 2\|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2\,\Re\big(\overline{\mathbf{x}}^2\mathbf{z}^2\big)\\
&\quad + 2|\mathbf{x}|^2\Big(\|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2|\mathbf{z}|^2 - \Re\Big(\big(\mathbf{A}\circ\overline{\mathbf{B}}\big)^H\mathbf{b}\Big)\Big)\\
&= \|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2\big(2|\mathbf{x}|^2|\mathbf{z}|^2 + 2\,\Re\big(\overline{\mathbf{x}}^2\mathbf{z}^2\big)\big)\\
&= \|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2(\overline{\mathbf{x}}\mathbf{z} + \mathbf{x}\overline{\mathbf{z}})^2 \geq 0
\end{aligned} \quad (80)$$

where the equality can be attained with $\mathbf{x} = \mathbf{z}_I - j\mathbf{z}_R$.

That $\nabla^2 f(\mathbf{z})$ is rank-deficient is implied by Theorem 2(3a).

(3b) This follows directly from Theorem 2(3b).

(3c) Let us perturb $\mathbf{z}$ by $\epsilon\in\mathbb{C}$ with $|\epsilon|$ being arbitrarily small and observe

$$\begin{aligned}
f(\mathbf{z}+\epsilon) &= \frac{1}{2}\|(\mathbf{A}(\mathbf{z}+\epsilon))\circ\big(\overline{(\mathbf{B}\mathbf{z}+\epsilon)}\big) - \mathbf{b}\|_2^2\\
&= \frac{1}{2}\|(\mathbf{A}\mathbf{z})\circ\big(\overline{\mathbf{B}\mathbf{z}}\big) + (\mathbf{A}\mathbf{z})\circ\big(\overline{\mathbf{B}\epsilon}\big)\\
&\quad + (\mathbf{A}\epsilon)\circ\big(\overline{\mathbf{B}\mathbf{z}}\big) + (\mathbf{A}\epsilon)\circ\big(\overline{\mathbf{B}\epsilon}\big) - \mathbf{b}\|_2^2\\
&= \frac{1}{2}\|(\mathbf{A}\mathbf{z})\circ\big(\overline{\mathbf{B}\mathbf{z}}\big) - \mathbf{b} + \mathbf{A}\circ\overline{\mathbf{B}}\big(\overline{\epsilon}\mathbf{z} + \epsilon\overline{\mathbf{z}} + |\epsilon|^2\big)\|_2^2\\
&= \frac{1}{2}\|(\mathbf{A}\mathbf{z})\circ\big(\overline{\mathbf{B}\mathbf{z}}\big) - \mathbf{b}\|_2^2\\
&\quad + \frac{1}{2}\|\mathbf{A}\circ\overline{\mathbf{B}}\big(2\,\Re(\overline{\epsilon}\mathbf{z}) + |\epsilon|^2\big)\|_2^2\\
&\quad + \Re\Big(\big((\mathbf{A}\mathbf{z})\circ\big(\overline{\mathbf{B}\mathbf{z}}\big) - \mathbf{b}\big)^H\\
&\quad \cdot\big(\mathbf{A}\circ\overline{\mathbf{B}}\big(2\,\Re(\overline{\epsilon}\mathbf{z}) + |\epsilon|^2\big)\big)\Big)\\
&= \frac{1}{2}\|(\mathbf{A}\mathbf{z})\circ\big(\overline{\mathbf{B}\mathbf{z}}\big) - \mathbf{b}\|_2^2\\
&\quad + \frac{1}{2}\|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2\big(2\,\Re(\overline{\epsilon}\mathbf{z}) + |\epsilon|^2\big)^2\\
&\quad + \Re\big(\|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2|\mathbf{z}|^2 - \mathbf{b}^H\big(\mathbf{A}\circ\overline{\mathbf{B}}\big)\big)\\
&\quad \cdot\big(2\,\Re(\overline{\epsilon}\mathbf{z}) + |\epsilon|^2\big)\\
&= f(\mathbf{z}) + \frac{1}{2}\|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2\big(2\,\Re(\overline{\epsilon}\mathbf{z}) + |\epsilon|^2\big)^2\\
&\quad + \Big(\|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2|\mathbf{z}|^2 - \Re\Big(\big(\mathbf{A}\circ\overline{\mathbf{B}}\big)^H\mathbf{b}\Big)\Big)\\
&\quad \cdot\big(2\,\Re(\overline{\epsilon}\mathbf{z}) + |\epsilon|^2\big)\\
&= f(\mathbf{z}) + \frac{1}{2}\|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2\big(2\,\Re(\overline{\epsilon}\mathbf{z}) + |\epsilon|^2\big)^2
\end{aligned} \quad (81)$$

where the last equality is given by Equation (79). As a result

$$f(\mathbf{z}+\epsilon) - f(\mathbf{z}) = \frac{1}{2}\|\mathbf{A}\circ\overline{\mathbf{B}}\|_2^2\big(2\,\Re(\overline{\epsilon}\mathbf{z}) + |\epsilon|^2\big)^2 \geq 0, \quad (82)$$

that is, $\mathbf{z}$ is a local minimum. □

ACKNOWLEDGMENT

TerraSAR-X data were provided by the German Aerospace Center (DLR) under the TerraSAR-X New Modes AO Project LAN2188.

The author N. Ge would like to thank Dr. S. Auer and Dr. C. Gisinger for preparing Level-0 ground-truth data, Dr. S. Li for the insight from an algebraic geometric point of view, F. Rodriguez Gonzalez and A. Parizzi for valuable discussions, Dr. H. Ansari and Dr. S. Montazeri for explaining the sequential estimator and EMI, Dr. F. De Zan for providing an overview of multi-look multi-master SAR tomography in the forested scenarios, Dr. Y. Huang and Prof. Dr. L. Ferro-Famil for discussing a preliminary version of this article, and the reviewers for their constructive and insightful comments.

REFERENCES

[1] A. Reigber and A. Moreira, "First demonstration of airborne SAR tomography using multibaseline L-band data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2142–2152, Apr. 2000.

[2] F. Gini, F. Lombardini, and M. Montanari, "Layover solution in multibaseline SAR interferometry," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 4, pp. 1344–1356, Oct. 2002.

[3] G. Fornaro, F. Serafino, and F. Soldovieri, "Three-dimensional focusing with multipass SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 507–517, Mar. 2003.

[4] F. Lombardini, "Differential tomography: A new framework for SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 37–44, Jan. 2005.

[5] G. Fornaro, D. Reale, and F. Serafino, "Four-dimensional SAR imaging for height estimation and monitoring of single and double scatterers," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 224–237, Jan. 2009.

[6] X. X. Zhu and R. Bamler, "Let's do the time warp: Multicomponent nonlinear motion estimation in differential SAR tomography," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 735–739, Jul. 2011.

[7] A. Ferretti, C. Prati, and F. Rocca, "Permanent scatterers in SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 1, pp. 8–20, May 2001.

[8] C. Colesanti, A. Ferretti, F. Novali, C. Prati, and F. Rocca, "SAR monitoring of progressive and seasonal ground deformation using the permanent scatterers technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 7, pp. 1685–1701, Jul. 2003.

[9] N. Adam, B. Kampes, and M. Eineder, "Development of a scientific permanent scatterer system: Modifications for mixed ERS/ENVISAT time series," in *Proc. Envisat Symp.*, vol. 572, 2005.

[10] B. M. Kampes, *Radar Interferometry: Persistent Scatterer Technique*, vol. 12. Dordrecht, The Netherlands: Springer, 2006.

[11] A. Budillon, A. Evangelista, and G. Schirinzi, "Three-dimensional SAR focusing from multipass signals using compressive sampling," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 488–499, Jan. 2011.

[12] X. X. Zhu and R. Bamler, "Tomographic SAR inversion by L_1-norm regularization—The compressive sensing approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3839–3846, Oct. 2010.

[13] E. Aguilera, M. Nannini, and A. Reigber, "Multisignal compressed sensing for polarimetric SAR tomography," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 5, pp. 871–875, Sep. 2012.

[14] M. Schmitt and U. Stilla, "Compressive sensing based layover separation in airborne single-pass multi-baseline InSAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 313–317, Mar. 2013.

[15] G. Fornaro, S. Verde, D. Reale, and A. Pauciullo, "CAESAR: An approach based on covariance matrix decomposition to improve multibaseline–multitemporal interferometric SAR processing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2050–2065, Apr. 2015.

[16] X. X. Zhu, N. Ge, and M. Shahzad, "Joint sparsity in SAR tomography for urban mapping," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 8, pp. 1498–1509, Dec. 2015.

[17] N. Ge, F. R. Gonzalez, Y. Wang, Y. Shi, and X. X. Zhu, "Spaceborne staring spotlight SAR tomography—A first demonstration with TerraSAR-X," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 10, pp. 3743–3756, Oct. 2018.

[18] Y. Shi, X. X. Zhu, W. Yin, and R. Bamler, "A fast and accurate basis pursuit denoising algorithm with application to super-resolving tomographic SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6148–6158, Oct. 2018.

[19] Y. Shi, Y. Wang, J. Kang, M. Lachaise, X. X. Zhu, and R. Bamler, "3D reconstruction from very small TanDEM-X stacks," in *Proc. 12th Eur. Conf. Synth. Aperture Radar*, 2018, pp. 1–4.

[20] Y. Shi, X. X. Zhu, and R. Bamler, "Nonlocal compressive sensing-based SAR tomography," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 3015–3024, May 2019.

[21] G. Fornaro, F. Lombardini, and F. Serafino, "Three-dimensional multipass SAR focusing: Experiments with long-term spaceborne data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 702–714, Apr. 2005.

[22] X. X. Zhu and R. Bamler, "Very high resolution spaceborne SAR tomography in urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4296–4308, Dec. 2010.

[23] X. X. Zhu and R. Bamler, "Super-resolution power and robustness of compressive sensing for spectral estimation with application to spaceborne tomographic SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 247–258, Jan. 2012.

[24] X. X. Zhu and R. Bamler, "Sparse tomographic SAR reconstruction from mixed TerraSAR-X/TanDEM-X data stacks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 7468–7471.

[25] N. Ge and X. X. Zhu, "Bistatic-like differential SAR tomography," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5883–5893, Aug. 2019.

[26] L. Liang, X. Li, L. Ferro-Famil, H. Guo, L. Zhang, and W. Wu, "Urban area tomography using a sparse representation based two-dimensional spectral analysis technique," *Remote Sens.*, vol. 10, no. 2, p. 109, Jan. 2018.

[27] D. Perissin and T. Wang, "Repeat-pass SAR interferometry with partially coherent targets," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 271–280, Jan. 2012.

[28] A. Ferretti, A. Fumagalli, F. Novali, C. Prati, F. Rocca, and A. Rucci, "A new algorithm for processing interferometric data-stacks: SqueeSAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 9, pp. 3460–3470, Sep. 2011.

[29] A. Parizzi and R. Brcic, "Adaptive InSAR stack multilooking exploiting amplitude statistics: A comparison between different techniques and practical results," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 441–445, May 2011.

[30] H. Ansari, F. De Zan, and R. Bamler, "Sequential estimator: Toward efficient InSAR time series analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5637–5652, Oct. 2017.

[31] H. Ansari, F. De Zan, and R. Bamler, "Efficient phase estimation for interferogram stacks," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 4109–4125, Jul. 2018.

[32] S. Duque, P. Lopez-Dekker, J. J. Mallorqui, A. Y. Nashashibi, and A. M. Patel, "Experimental results with bistatic SAR tomography," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2009, p. 37.

[33] S. Duque, P. Lopez-Dekker, J. C. Merlano, and J. J. Mallorqui, "Bistatic SAR tomography: Processing and experimental results," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2010, pp. 154–157.

[34] S. Duque, C. Rossi, and T. Fritz, "Single-pass tomography with alternating bistatic TanDEM-X data," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 409–413, Feb. 2015.

[35] S. R. Cloude, "Polarization coherence tomography," *Radio Sci.*, vol. 41, no. 4, p. 506, 2006.

[36] F. De Zan, K. Papathanassiou, and S. Lee, "Tandem-L forest parameter performance analysis," in *Proc. Int. Workshop Appl. Polarimetry Polarimetric Interferometry*, Frascati, Italy, 2009, pp. 1–6.

[37] S. Tebaldini, "Algebraic synthesis of forest scenarios from multibaseline PolInSAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 12, pp. 4132–4142, Dec. 2009.

[38] S. Tebaldini, "Single and multipolarimetric SAR tomography of forested areas: A parametric approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2375–2387, May 2010.

[39] Y. Huang, L. Ferro-Famil, and A. Reigber, "Under-foliage object imaging using SAR tomography and polarimetric spectral estimators," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2213–2225, Jun. 2012.

[40] S.-K. Lee and T. E. Fatoyinbo, "TanDEM-X pol-InSAR inversion for mangrove canopy height estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3608–3618, Jul. 2015.

[41] M. Pardini and K. Papathanassiou, "On the estimation of ground and volume polarimetric covariances in forest scenarios with SAR tomography," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1860–1864, Oct. 2017.

[42] V. Cazcarra-Bes, M. Pardini, M. Tello, and K. P. Papathanassiou, "Comparison of tomographic SAR reflectivity reconstruction algorithms for forest applications at L-band," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 147–164, Jan. 2020.

[43] G. Krieger *et al.*, "TanDEM-X: A satellite formation for high-resolution SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3317–3341, Nov. 2007.

[44] K. Goel and N. Adam, "Fusion of monostatic/bistatic InSAR stacks for urban area analysis via distributed scatterers," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 4, pp. 733–737, Apr. 2014.

[45] P. Berardino, G. Fornaro, R. Lanari, and E. Sansosti, "A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2375–2383, Nov. 2002.

[46] R. Lanari, O. Mora, M. Manunta, J. J. Mallorqui, P. Berardino, and E. Sansosti, "A small-baseline approach for investigating deformations on full-resolution differential SAR interferograms," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1377–1386, Jul. 2004.

[47] A. Hooper, "A multi-temporal InSAR method incorporating both persistent scatterer and small baseline approaches," *Geophys. Res. Lett.*, vol. 35, no. 16, 2008.

[48] A. Moreira *et al.*, "Tandem-L: A highly innovative bistatic SAR mission for global observation of dynamic processes on the Earth's surface," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 2, pp. 8–23, Jun. 2015.

[49] J. A. Bondy and U. S. R. Murty, *Graph Theory*. Berlin, Germany: Springer, 2008.

[50] I. Hajnsek, T. Busche, G. Krieger, M. Zink, and A. Moreira, *Announcement of Opportunity: TanDEM-X Science Phase*, document TD-PD-PL-0032, 2014.

[51] P. Stoica and Y. Selen, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.

[52] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.

[53] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[54] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. New York, NY, USA: Springer, 2017.

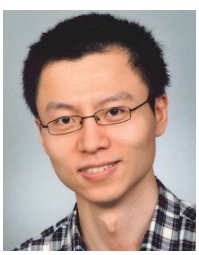
[55] J. Nocedal and S. Wright, *Numerical Optimization*. New York, NY, USA: Springer, 2006.

[56] J. Sun, Q. Qu, and J. Wright, "When are nonconvex problems not scary?" 2015, *arXiv:1510.06096*. [Online]. Available: http://arxiv.org/abs/1510.06096

[57] E. Freitag and R. Busam, *Complex Analysis*. Berlin, Germany: Springer, 2009.

[58] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[59] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "CoSpace: Common subspace learning from hyperspectral-multispectral correspondences," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4349–4359, Jul. 2019.
[60] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. X. Zhu, "Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 147, pp. 193–205, Jan. 2019.
[61] N. Parikh *et al.*, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.
[62] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
[63] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[64] M. Fornasier, S. Peter, H. Rauhut, and S. Worm, "Conjugate gradient acceleration of iteratively re-weighted least squares methods," *Comput. Optim. Appl.*, vol. 65, no. 1, pp. 205–259, Sep. 2016.
[65] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, May 2011.
[66] N. Adam, F. R. Gonzalez, A. Parizzi, and R. Brcic, "Wide area persistent scatterer interferometry: Current developments, algorithms and examples," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2013, pp. 1857–1860.
[67] F. R. Gonzalez, N. Adam, A. Parizzi, and R. Brcic, "The integrated wide area processor (IWAP): A processor for wide area persistent scatterer interferometry," in *Proc. ESA Living Planet Symp.*, vol. 722, 2013, p. 353.
[68] S. Auer, C. Gisinger, N. Ge, F. R. Gonzalez, and X. X. Zhu, "SIGS: A SAR imaging geodesy and simulation framework for generating 3-D ground truth," German Aerosp. Center (DLR), Wessling, Germany, 2020.
[69] J. Revels, M. Lubin, and T. Papamarkou, "Forward-mode automatic differentiation in Julia," 2016, *arXiv:1607.07892*. [Online]. Available: http://arxiv.org/abs/1607.07892
[70] P. K Mogensen and A. N Riseth, "Optim: A mathematical optimization package for Julia," *J. Open Source Softw.*, vol. 3, no. 24, p. 615, Apr. 2018.
[71] S. J. Auer, "3D synthetic aperture radar simulation for interpreting complex urban reflection scenarios," Ph.D. dissertation, Dept. Bauingenieur- und Vermessungswesen, Technische Univ. München, München, Germany, 2011.
[72] M. Eineder, C. Minet, P. Steigenberger, X. Cong, and T. Fritz, "Imaging geodesy—Toward centimeter-level ranging accuracy with TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 661–671, Oct. 2010.
[73] C. Gisinger *et al.*, "Precise three-dimensional stereo localization of corner reflectors and persistent scatterers with TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1782–1802, Apr. 2015.
[74] S. Hackel, C. Gisinger, U. Balss, M. Wermuth, and O. Montenbruck, "Long-term validation of TerraSAR-X and TanDEM-X orbit solutions with laser and radar measurements," *Remote Sens.*, vol. 10, no. 5, p. 762, May 2018.
[75] S. Gernhardt, S. Auer, and K. Eder, "Persistent scatterers at building facades—Evaluation of appearance and localization accuracy," *ISPRS J. Photogramm. Remote Sens.*, vol. 100, pp. 92–105, Feb. 2015.



Nan Ge received the bachelor's and master's degrees in geoenvironmental engineering from the Clausthal University of Technology, Clausthal-Zellerfeld, Germany, and the Ph.D. degree in signal processing from the Technical University of Munich, Munich, Germany.

In 2013, he joined the German Aerospace Center, Wessling, Germany, where he has been developing advanced tomographic SAR techniques with Remote Sensing Technology Institute.

Richard Bamler (Fellow, IEEE) received the diploma degree in electrical engineering, the Ph.D. degree in engineering, and the Habilitation degree in signal and systems theory from the Technical University of Munich (TUM), Munich, Germany, in 1980, 1986, and 1988, respectively.

He was with the Technical University of Munich from 1981 to 1989 on optical signal processing, holography, wave propagation, and tomography. In 1989, he joined the German Aerospace Center (DLR), Wessling, Germany,, where he is the Director of the Remote Sensing Technology Institute. In early 1994, he was a Visiting Scientist with the Jet Propulsion Laboratory (JPL) in preparation of the SIC-C/X-SAR missions, and in 1996, he was a Guest Professor with the University of Innsbruck, Innsbruck, Austria. Since 2003, he has held a Full Professorship in remote sensing technology with TUM as a double appointment with his DLR position. His teaching activities include university lectures and courses on signal processing, estimation theory, and SAR. He, his team, and his institute have been working on SAR and optical remote sensing, image analysis and understanding, machine learning, stereo reconstruction, computer vision, ocean color, hyperspectral image analysis, passive and active atmospheric sounding, and laboratory spectrometry. They were and are responsible for the development of the operational processors for SIR-C/X-SAR, SRTM, TerraSAR-X, TanDEM-X, PAZ, Tandem-L, ERS-2/GOME, ENVISAT/SCIAMACHY, MetOp/GOME-2, Sentinel 5p, Sentinel 4, DESIS, EnMAP, and so on. His research interests are in algorithms for optimum information extraction from remote sensing data. This involves new estimation algorithms, like sparse reconstruction and compressive sensing for SAR, multi/hyperspectral imaging, data fusion, and machine learning for Big EO Data analytics.

Danfeng Hong (Member, IEEE) received the M.Sc. degree *(summa cum laude)* in computer vision from the College of Information Engineering, Qingdao University, Qingdao, China, in 2015, and the Dr.-Ing. degree *(summa cum laude)* from Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), Munich, Germany, in 2019.

Since 2015, he also worked as a Research Associate with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany. He is a Research Scientist and leads a Spectral Vision Group at EO Data Science, IMF, DLR, and also an Adjunct Scientist with the GIPSA-lab, Grenoble INP, CNRS, University Grenoble Alpes, Grenoble, France. His research interests include signal/image processing and analysis, pattern recognition, machine/deep learning, and their applications in Earth Vision.

Xiao Xiang Zhu (Senior Member, IEEE) received the M.Sc., Dr.-Ing., and the Habilitation degrees in signal processing from the Technical University of Munich (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

Since 2019, she has been a Co-Coordinator of the Munich Data Science Research School (www.mu-ds.de). Since 2019, she also heads the Helmholtz Artificial Intelligence Cooperation Unit (HAICU)—Research Field Aeronautics, Space and Transport. Since May 2020, she has been the Director of the International Future AI Lab "AI4EO—Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond," Munich, Germany (ai4eo.de). She was a Guest Scientist or Visiting Professor with the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China, the University of Tokyo, Tokyo, Japan, and University of California at Los Angeles, Los Angeles, CA, USA, in 2009, 2014, 2015, and 2016, respectively. She is the Professor for Signal Processing in Earth Observation (www.sipeo.bgu.tum.de), Technical University of Munich (TUM), and the Head of the Department EO Data Science, Remote Sensing Technology Institute, German Aerospace Center (DLR), Wessling, Germany. Her main research interests are remote sensing and Earth observation, signal processing, machine learning, and data science, with a special application focus on global urban mapping.

Dr. Zhu is a member of Young Academy (Junge Akademie/Junges Kolleg) at the Berlin-Brandenburg Academy of Sciences and Humanities, the German National Academy of Sciences Leopoldina, and the Bavarian Academy of Sciences and Humanities. She is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.